

(19) 日本国特許庁(JP)

(12) 公表特許公報(A)

(11) 特許出願公表番号

特表2004-534519

(P2004-534519A)

(43) 公表日 平成16年11月18日(2004.11.18)

(51) Int. Cl. ⁷	F I	テーマコード (参考)
C12N 15/09	C12N 15/00 A	2G045
C07K 1/22	C07K 1/22	4B024
C12Q 1/02	C12Q 1/02	4B063
G01N 27/62	G01N 27/62 D	4C086
G01N 33/15	G01N 33/15 Z	4C206
審査請求 未請求 予備審査請求 有 (全 205 頁) 最終頁に続く		

(21) 出願番号 特願2002-558871 (P2002-558871)
 (86) (22) 出願日 平成13年11月19日 (2001.11.19)
 (85) 翻訳文提出日 平成15年5月19日 (2003.5.19)
 (86) 国際出願番号 PCT/US2001/043348
 (87) 国際公開番号 W02002/058533
 (87) 国際公開日 平成14年8月1日 (2002.8.1)
 (31) 優先権主張番号 60/249,832
 (32) 優先日 平成12年11月17日 (2000.11.17)
 (33) 優先権主張国 米国 (US)
 (31) 優先権主張番号 60/329,463
 (32) 優先日 平成13年10月15日 (2001.10.15)
 (33) 優先権主張国 米国 (US)

(71) 出願人 503181196
 スランツ アルフレッド イー.
 アメリカ合衆国 マサチューセッツ州 コ
 ハセット ニコラス ロード 14
 (74) 代理人 100102978
 弁理士 清水 初志
 (74) 代理人 100108774
 弁理士 橋本 一憲
 (72) 発明者 スランツ アルフレッド イー.
 アメリカ合衆国 マサチューセッツ州 コ
 ハセット ニコラス ロード 14
 Fターム(参考) 2G045 AA34 AA35 AA40 BA11 BB50
 DA13 DA36 FB02
 4B024 AA01 AA11 CA04 HA14

最終頁に続く

(54) 【発明の名称】 標的分子機能の決定と薬物リード化合物の同定に関する方法

(57) 【要約】

本発明は、化学リガンドを標的とする機能の決定および薬物リード化合物の同定に使用する方法に関する。

BEST AVAILABLE COPY

【特許請求の範囲】

【請求項1】

標的分子に結合する候補リガンドを選択する方法であって、

(a) 標的分子を含むインビトロ試料と候補リガンドのライブラリとを、該標的分子と1つ以上の該候補リガンドとの間で複合体が形成され得る条件下で接触させる段階（ここで該ライブラリは少なくとも2つの異なる化学的骨格または少なくとも11種類の化合物を含む）、

(b) 前記複合体を単離する段階、

(c) 1つ以上の前記候補リガンドを前記複合体から回収する段階、および

(d) 1つ以上の回収した候補リガンドを同定する段階

を含む方法。

10

【請求項2】

段階(d)が、前記の回収した候補リガンドのMS、IR、FTIR、NMR、および／またはUVスペクトルを測定することを含む、請求項1記載の方法。

【請求項3】

少なくとも100個の異なる候補リガンドを同時に前記標的分子と接触させる、請求項1記載の方法。

【請求項4】

標的分子に結合する候補リガンドを選択する方法であって、

(a) 標的分子を含むインビトロ試料と候補リガンドのライブラリとを、該標的分子と1つ以上の該候補リガンドとの間で複合体が形成され得る条件下で接触させる段階、

(b) 前記複合体を単離する段階、

(c) 1つ以上の前記候補リガンドを前記複合体から回収する段階、および

(d) 回収した候補リガンドの質量スペクトル中の同位体ピークまたはフラグメントピークの質量／電荷比を決定する段階により、該回収候補リガンドを同定する段階を含む方法。

20

【請求項5】

少なくとも100個の異なる候補リガンドを同時に前記標的分子と接触させる、請求項4記載の方法。

【請求項6】

段階(d)が、前記回収候補リガンドの質量スペクトル中の親ピークの質量／電荷比を決定する段階を更に含む、請求項4記載の方法。

30

【請求項7】

標的分子に結合する候補リガンドを選択する方法であって、

(a) 未知の生物学的機能を持つ標的分子を含むインビトロ試料と候補リガンドのライブラリとを、該標的分子と1つ以上の該候補リガンドとの間で複合体が形成され得る条件下で接触させる段階、

(b) 前記複合体を単離する段階、

(c) 1つ以上の前記候補リガンドを前記複合体から回収する段階、ならびに

(d) 回収した候補リガンドのMS、IR、FTIR、NMR、および／またはUVスペクトルを測定することにより、該回収候補リガンドを同定する段階を含む方法。

40

【請求項8】

少なくとも100個の異なる候補リガンドを同時に前記標的分子と接触させる、請求項7記載の方法。

【請求項9】

標的分子に結合する候補リガンドを選択する方法であって、

(a) 標的分子を含むインビトロ試料と1つ以上の候補リガンドとを、該標的分子と1つ以上の該候補リガンドとの間で複合体が形成され得る条件下で接触させる段階、

(b) 前記複合体を単離する段階、

50

(c) 1つ以上の前記候補リガンドを前記複合体から回収する段階、ならびに
(d) 回収した候補リガンドのIR、FTIR、NMR、および/またはUVスペクトルを測定することにより、該回収候補リガンドを同定する段階を含む方法。

【請求項10】

少なくとも100個の異なる候補リガンドを同時に前記標的分子と接触させる、請求項9記載の方法。

【請求項11】

標的分子に結合する候補リガンドを選択する方法であって、

- (a) 第一標的分子および第二標的分子を含むインビトロ試料と候補リガンドのライブラリとを、該第一標的分子と1つ以上の該候補リガンドとの間で複合体および該第二標的分子と1つ以上の該候補リガンドとの間で複合体が形成され得る条件下で接触させる段階、
(b) 候補リガンドに結合した前記第一標的分子を含む第一複合体を単離する段階、および候補リガンドに結合した前記第二標的分子を含む第二複合体を単離する段階、
(c) 1つ以上の前記候補リガンドを前記第一複合体から、および/または前記第二複合体から回収する段階、ならびに
(d) 1つ以上の回収した候補リガンドを同定する段階を含む方法。

10

【請求項12】

前記試料を、前記標的分子、前記第一標的分子、または前記第二標的分子と結合することが知られている競合リガンドと接触させる段階を更に含む、請求項11記載の方法。

20

【請求項13】

標的分子の生物学的機能を決定する方法であって、

- (a) 未知の生物学的機能を持つ標的分子を含むインビトロ試料と候補リガンドのライブラリとを、1つ以上の該候補リガンドが該標的分子に結合し得る条件下で接触させる段階、
(b) 前記標的分子と結合する候補リガンドを選択する段階、および
(c) 選択した候補リガンドの生物学的アッセイにおける作用を測定することにより、前記標的分子の生物学的機能を決定する段階を含む方法。

30

【請求項14】

前記の選択した候補リガンドを同定する段階を更に含む、請求項13記載の方法。

【請求項15】

標的分子の生物学的機能を決定する方法であって、

- (a) 疾病状態、生理学的刺激物質の存在下、または特定の細胞的もしくは生物学的過程において上方制御または下方制御される標的分子を含むインビトロ試料と候補リガンドのライブラリとを、1つ以上の該候補リガンドが該標的分子に結合し得る条件下で接触させる段階、
(b) 前記標的分子と結合する候補リガンドを選択する段階、および
(c) 選択した候補リガンドの生物学的アッセイにおける作用を測定することにより、前記標的分子の生物学的機能を決定する段階を含む方法。

40

【請求項16】

前記の選択した候補リガンドを同定する段階を更に含む、請求項15記載の方法。

【請求項17】

前記の選択した候補リガンドが前記生物学的アッセイにおける前記標的分子の活性を増大させる、請求項15記載の方法。

【請求項18】

前記の選択した候補リガンドが前記生物学的アッセイにおける前記標的分子の活性を低下させる、請求項15記載の方法。

50

【請求項 19】

標的分子の生物学的機能を決定する方法であって、

- (a) 標的分子を含むインビトロ試料と候補リガンドのライブラリとを、1つ以上の該候補リガンドが該標的分子に結合し得る条件下で接触させる段階、
- (b) 前記標的分子と結合した候補リガンドを選択する段階、および
- (c) 選択した前記候補リガンドが、疾病もしくは障害を有するかまたは生理学的刺激物質の存在下もしくは非存在下で特定の細胞的もしくは生物学的過程にある生物体の組織に及ぼす効果を測定することにより、前記標的分子の生物学的機能を決定する段階を含む方法。

【請求項 20】

前記組織がヒトの組織である、請求項19記載の方法。

【請求項 21】

関心対象の標的分子に結合する2つのリガンドを反応させる方法であって、未知の二次構造または三次構造を有する標的分子を含む細胞またはインビトロ試料と、第一架橋剤を含む第一リガンドおよび第二リガンドとを、該標的分子が該第一リガンドおよび第二リガンドと結合しかつ該第一架橋剤が該第二リガンドと共有結合し得る条件下で接触させる段階により、該第一リガンドと該第二リガンドとを含む架橋生成物を形成させる段階を含む方法。

【請求項 22】

関心対象の標的分子に結合する2つのリガンドを反応させる方法であって、標的分子を含む細胞またはインビトロ試料と、第一架橋剤を含む第一リガンドおよび第二リガンドとを接触（ここで該標的分子における該第一リガンドまたは該第二リガンドへの結合部位の位置または三次構造は未知であり、かつ該接触段階が、該標的分子が該第一リガンドおよび第二リガンドと結合しかつ該第一架橋剤が第二リガンドと共有結合し得る条件下で実施される）させることにより、該第一リガンドと該第二リガンドとを含む架橋生成物を形成させる段階を含む方法。

【請求項 23】

関心対象の標的分子に結合する2つのリガンドを反応させる方法であって、標的分子を含む細胞またはインビトロ試料と、第一架橋剤を含む第一リガンドおよび第二リガンドとを接触（ここで該接触段階が、該標的分子が該第一リガンドおよび該第二リガンドと結合しかつ該第一架橋剤が該第二リガンドと共有結合し得る条件下で実施される）させることにより、該第一リガンドおよび該第二リガンドを含みかつ該標的分子に対して該第一リガンドまたは該第二リガンドより高い親和性を有する架橋生成物を形成させる段階を含む方法。

【請求項 24】

異なる標的分子に結合する2つのリガンドを反応させる方法であって、第一標的分子および第二標的分子を含む細胞またはインビトロ試料と、第一架橋剤を含む第一リガンドおよび第二リガンドとを接触（ここで該接触段階が、

- (i) 該第一タンパク質の該第一リガンドへの結合、
- (ii) 該第二タンパク質の該第二リガンドへの結合、および
- (iii) 該第一架橋剤の該第二リガンドへの共有結合

を可能にする条件下で実施される）させることにより、該第一リガンドと該第二リガンドとを含む架橋生成物を形成させる段階を含み、ここで、該第一標的分子における該第一リガンドへの結合部位の位置もしくは三次構造および／または該第二標的分子における該第二リガンドへの結合部位の位置もしくは三次構造は未知である方法。

【請求項 25】

前記架橋生成物の形成が、前記第一タンパク質および前記第二タンパク質とがインビボで相互作用することを示す、請求項24記載の方法。

【請求項 26】

第一タンパク質に結合する第二タンパク質を単離する方法であって、

10

20

30

40

50

(a) 第一タンパク質および第二タンパク質を含む細胞またはインビトロ試料と、第一架橋剤を含む第一リガンドおよび第二リガンドとを、

(i) 該第一タンパク質の該第一リガンドへの結合、

(ii) 該第二タンパク質の該第二リガンドへの結合、および

(iii) 該第一架橋剤の該第二リガンドへの共有結合

を可能にする条件下で接触させる段階により、該第一リガンドと該第二リガンドとを含む架橋生成物、ならびに該架橋生成物、該第一タンパク質、および該第二タンパク質を含む複合体を形成させる段階、

(b) 前記複合体を単離する段階、ならびに

(c) 前記複合体中で、または前記複合体から回収した、前記第一タンパク質および／または前記第二タンパク質を同定する段階を含む方法。

10

【請求項 27】

前記第一タンパク質が検出可能な基を含む、請求項26記載の方法。

【請求項 28】

前記第二リガンドが架橋剤を含む、請求項26記載の方法。

【請求項 29】

前記架橋生成物の形成が、前記第一タンパク質と前記第二タンパク質とがインビボで相互作用することを意味する、請求項26記載の方法。

【請求項 30】

前記架橋生成物の前記標的分子に対する親和性が、前記第一リガンドまたは前記第二リガンドの前記標的分子に対する親和性よりも高い、請求項26記載の方法。

20

【請求項 31】

前記架橋生成物が、薬物の発見もしくは開発またはリード化合物の最適化において使用される、請求項26記載の方法。

【請求項 32】

前記架橋生成物が、農業用または環境用の物質の開発に使用される、請求項26記載の方法。

【請求項 33】

関心対象の小分子と結合する候補標的分子を選択する方法であって、

30

(a) アミノ酸以外の部分を持つまたは4000ダルトン未満の分子量を持つ関心対象の小分子を含むインビトロ試料と、標的分子候補のライブラリとを、該関心対象の小分子と1つ以上の該標的分子候補との間で複合体が形成され得る条件下で接触させる段階（ここで該標的分子はフェージ表面で発現されない）、

(b) 前記複合体を単離する段階、および

(c) 1つ以上の前記標的分子候補を前記複合体から回収する段階により、前記の関心対象の小分子に結合する1つ以上の標的分子候補を選択する段階を含む方法。

【請求項 34】

段階 (a) に先立ち、前記の関心対象の小分子を、生物学的アッセイ中のその作用に基づいて小分子ライブラリから選択する、請求項33記載の方法。

40

【請求項 35】

関心対象の小分子に結合する標的タンパク質を選択する方法であって、

(a) 表面タンパク質に共有結合した標的タンパク質を含むタンパク質融合体を細胞集団内で発現させる段階（ここで該発現段階は該タンパク質融合体が該細胞表面にディスプレイされ得る条件下で実施する）、

(b) 前記細胞を、アミノ酸以外の部分を持つまたは4000ダルトン未満の分子量を有する、関心対象の小分子と接触させる段階、および

(c) 前記の関心対象の小分子と結合した前記細胞を選択する段階により、前記の関心対象の小分子に結合する前記標的タンパク質を選択する段階

50

を含む方法。

【請求項36】

前記細胞が哺乳動物細胞、細菌細胞、酵母細胞または昆虫細胞である、請求項35記載の方法。

【請求項37】

関心対象の小分子に結合する標的タンパク質を選択する方法であって、

(a) 表面タンパク質に共有結合した標的タンパク質を含むタンパク質融合体を細胞集団内で発現させる段階（ここで該発現段階は、該タンパク質融合体がウイルスに感染した該細胞から遊離した該ウイルスの表面上でディスプレイされ得る条件下で実施する）、

(b) 前記ウイルスを関心対象の小分子と接触させる段階（ここで、関心対象の小分子とは、

(i) 核酸である、

(ii) 糖質である、

(iii) 脂質である、

(iv) アミノ酸以外の部分を有する、

(v) 750ダルトン未満の分子量を有する、または

(vi) 細菌により天然に生成されない分子ではない

）、および

(c) 前記の関心対象の小分子と結合した前記ウイルスを選択する段階により、前記の関心対象の小分子に結合する前記標的タンパク質を選択する段階

を含む方法。

【請求項38】

前記ウイルスがバクテリオファージまたはアデノウイルスである、請求項37記載の方法。

【請求項39】

関心対象の小分子に結合する標的タンパク質を選択する方法であって、

(a) 標的タンパク質のライブラリを細胞集団またはインビトロ試料内で発現させる段階（ここで各標的タンパク質は、標的タンパク質をコードする核酸と共有結合している）、

(b) 前記細胞またはインビトロ試料と、アミノ酸以外の部分を有するかまたは4000ダルトン未満の分子量を有する関心対象の小分子とを接触させる段階、および

(c) 前記の関心対象の小分子と結合した前記の標的タンパク質を選択する段階

を含む方法。

【請求項40】

前記の選択された標的タンパク質を同定する段階を更に含む、請求項39記載の方法。

【請求項41】

少なくとも100個のヒト標的タンパク質を前記の関心対象の小分子に接触させる、請求項39記載の方法。

【請求項42】

前記の関心対象の小分子が天然に存在しない分子である、請求項39記載の方法。

【請求項43】

該標的分子を薬物標的として確認する前に、該標的分子と結合するかまたはその活性を調節する候補化合物を選択する方法であって、

(a) 以前に薬物標的として確認されていない標的分子を含む細胞またはインビトロ試料と候補化合物のライブラリとを、1つ以上の該候補化合物が該標的分子に結合するかまたはその活性を調節し得る条件下で接触させる段階、および

(b) 前記標的分子と結合するかまたはその活性を調節する候補化合物を選択する段階を含む方法。

【請求項44】

前記ライブラリが少なくとも5個の候補化合物を含む、請求項43記載の方法。

【請求項45】

(c) 前記の選択された候補化合物の作用を生物学的アッセイにおいて測定することによ

10

20

30

40

50

り前記標的分子の生物学的機能を決定する段階を更に含む、請求項43記載の方法。

【請求項46】

標的分子と結合するかまたはその活性を調節する候補化合物を選択する方法であって、

(a) 第一標的分子および第二標的分子を含む細胞またはインビトロ試料と候補化合物のライブラリとを、1つ以上の該候補化合物が該第一標的分子と結合するかまたはその活性を調節しかつ1つ以上の該候補化合物が該第二標的分子と結合するかまたはその活性を調節し得る条件下で接触させる段階、

(b) 該第一標的分子と結合するかまたはその活性を調節する候補化合物を選択する段階、および

(c) 該第二標的分子と結合するかまたはその活性を調節する候補化合物を選択する段階を含む方法。 10

【請求項47】

前記細胞またはインビトロ試料が少なくとも5個の標的分子を含み、かつ該標的分子各々に対して、該標的分子に結合するかまたはその活性を調節する候補化合物が選択される、請求項46記載の方法。

【請求項48】

前記標的分子に結合するかまたはその活性を調節するリガンドおよびその能力の記録に関連する該標的分子の記録を少なくとも10件含む、電子データベース。

【請求項49】

生物体プロテオーム中の少なくとも0.5%のタンパク質の記録を含む、請求項48記載のデータベース。 20

【請求項50】

標的ドメインに結合するリガンドおよびその能力の記録に関連する該標的分子ドメインの記録を少なくとも10件含む、電子データベース。

【請求項51】

前記標的分子に結合するかまたはその活性を調節するリガンドおよびその能力の記録に関連する、薬物標的として以前に確認されていない多数の該標的分子の記録を含む、電子データベース。

【請求項52】

請求項48、50、または51記載のデータベース、および、(i) コンピュータにその記録が保存されている標的分子に結合するかもしくはその活性を調節する1つ以上のリガンドを表示できる、または(ii) 該コンピュータにその記録が保存されているリガンドに結合するかもしくは該リガンドにより調節される活性を有する1つ以上の標的分子を表示できるユーザーインターフェースを含む、コンピュータ。 30

【請求項53】

前記化合物の影響を受ける1つ以上の生物学的アッセイにおける1つの表現型の記録に関連する、少なくとも1000件の該化合物の記録を含む電子データベースであって、該生物学的アッセイが該化合物に結合するタンパク質をコードする核酸の外因性コピーを含まない細胞またはインビトロ試料を含む、電子データベース。

【請求項54】

請求項53記載のデータベース、および、(i) コンピュータにその記録が保存されている化合物に対する1つ以上の生物学的アッセイにおける1つ以上の表現型を表示できる、または(ii) 該コンピュータにその記録が保存されている表現型に作用する1つ以上の化合物を表示できるユーザーインターフェースを含む、コンピュータ。 40

【請求項55】

標的分子の発現プロフィールまたは該標的分子の活性の記録に関連する、少なくとも10件の該標的分子の記録を含む、電子データベース。

【請求項56】

標的分子の発現プロフィールまたは該標的分子の活性の記録に関連する、薬物標的として以前に確認されていない多数の該標的分子の記録を含む、電子データベース。 50

【請求項 57】

請求項55または56記載のデータベース、および、(i) 前記コンピュータにその記録が保存されている標的分子の1つ以上の発現プロフィールもしくは活性の記録を表示できる、または(ii) 該コンピュータにその記録が保存されている発現プロフィールもしくは活性を有する1つ以上の標的分子を表示できるユーザーインターフェースを含む、コンピュータ。

【請求項 58】

関心対象の表現型に關与する標的分子を同定する方法であって、

(a) リガンドおよび前記表現型に寄与するその能力の記録に關連する、生物学的アッセイにおける多数の該表現型の記録を含む第一の電子データベースを提供する段階、

10

(b) 関心対象の表現型の選択を受ける段階、

(c) 前記第一データベースにおいて、前記の関心対象の表現型を生じる1つ以上のリガンドを同定する段階、

(d) 前記リガンドに結合するかまたはその活性が該リガンドにより調節される標的分子の記録に關連する、多数の該リガンドの記録を含む第二電子データベースを提供する段階、および

(e) 前記第二データベースにおいて、前記の関心対象の表現型を生じる前記のリガンドに結合するかまたは該リガンドにより調節される1つ以上の標的分子を同定する段階により、関心対象の該表現型に關与する1つ以上の標的分子を同定する段階

を含む方法。

20

【請求項 59】

前記の関心対象の表現型が疾病状態に關連しており、前記標的分子が該疾病状態を促進するかまたは抑制するかが決定される、請求項58記載の方法。

【請求項 60】

コンピュータで実行される、請求項58記載の方法。

【請求項 61】

関心対象の標的分子に關連する表現型を同定する方法であって、

(a) リガンドおよび前記標的分子に結合するかまたはその活性を調節するその能力の記録に關連する、多数の該標的分子の記録を含む第一電子データベースを提供する段階、

(b) 関心対象の標的分子の選択を受ける段階、

30

(c) 前記第一データベースにおいて関心対象の前記の標的分子に結合するかその活性を調節する1つ以上のリガンドを同定する段階、

(d) 前記リガンドにより生じる生物学的アッセイ中の表現型の記録に關連する、多数の該リガンドの記録を含む第二電子データベースを提供する段階、ならびに

(e) 前記第二データベースにおいて前記リガンドにより生じる1つ以上の表現型を同定する段階により、関心対象の前記標的分子に關連する1つ以上の表現型を同定する段階を含む方法。

【請求項 62】

コンピュータで実行される、請求項61記載の方法。

【請求項 63】

関心対象の標的分子に結合するかまたはその活性を調節するリガンドを同定する方法であって、

40

(a) リガンドおよび前記標的分子に結合するかまたはその活性を調節するその能力の記録に關連する、少なくとも10件の該標的分子の記録を含む電子データベースを提供する段階、

(b) 関心対象の標的分子の選択を受ける段階、ならびに

(c) 前記データベースにおいて関心対象の前記標的分子に結合するかまたはその活性を調節する1つ以上のリガンドを同定する段階を含む方法。

【請求項 64】

50

前記リガンドを薬物の発見もしくは開発またはリード化合物の最適化に使用する、請求項63記載の方法。

【請求項65】

前記リガンドを農業用または環境用の物質の開発に使用する、請求項63記載の方法。

【請求項66】

コンピュータで実行される、請求項63記載の方法。

【請求項67】

関心対象の前記標的分子に結合するかまたはその活性を調節する2つ以上のリガンドの化学構造を比較することにより、関心対象の該標的分子の結合または調節を促進する該リガンド中の官能基を同定する段階を更に含む、請求項63記載の方法。

10

【請求項68】

関心対象の前記標的分子に結合するかまたはその活性を調節する2つ以上のリガンドの化学構造を比較することにより、該リガンド集団における1つ以上の官能基または骨格の頻度を決定する段階を更に含む、請求項63記載の方法。

【請求項69】

2つ以上の前記リガンド中に存在する1つ以上の官能基を有する化合物を1つ以上作成する段階を更に含む、ここで該化合物が薬物の発見もしくは開発またはリード化合物の最適化に使用される、請求項63記載の方法。

【請求項70】

関心対象のリガンドに結合するかまたは該リガンドにより調節される活性を有する標的分子を同定する方法であって、

20

(a) 前記リガンドに結合するかまたは該リガンドにより調節される活性を有する標的分子の記録に関連する、少なくとも10件の該リガンドの記録を含む電子データベースを提供する段階、

(b) 関心対象のリガンドの選択を受ける段階、および

(c) 前記データベースにおいて関心対象の前記リガンドに結合するかまたは該リガンドにより調節される活性を有する1つ以上の標的分子を同定する段階を含む方法。

【請求項71】

コンピュータで実行される、請求項70記載の方法。

30

【請求項72】

関心対象のリガンドの選択性を決定する方法であって、

(a) リガンドおよび前記標的分子に結合するかまたはその活性を調節するその能力の記録に関連する、少なくとも10件の該標的分子の記録を含む電子データベースを提供する段階、

(b) 関心対象のリガンドの選択を受ける段階、ならびに

(c) 前記データベースにおいて前記リガンドに結合するかまたは該リガンドにより調節される標的分子数を決定する段階により、関心対象の該リガンドの選択性を決定する段階を含む方法。

【請求項73】

40

コンピュータで実行される、請求項72記載の方法。

【請求項74】

前記リガンドが標的分子の活性を増大させる請求項72記載の方法であって、該活性が疾病状態、有害な副作用、または毒性に関与しており、かつ該リガンドが薬物の発見もしくは開発またはリード化合物の最適化より除去される方法。

【請求項75】

前記リガンドが標的分子の活性を低減させる請求項72記載の方法であって、該活性が疾病状態、有害な副作用、または毒性に関与しており、かつ該リガンドが薬物の発見もしくは開発またはリード化合物の最適化の為に選択される方法。

【請求項76】

50

被験者の疾病または障害の治療、安定化、または予防のための治療法を選択する方法であって、

(a) 治療薬および前記標的分子に結合するかまたはその活性を調節するその能力の記録に関連する、少なくとも10件の該標的分子の記録を含む電子データベースを提供する段階

(b) 被験者において前記疾病または障害に関与する突然変異を有する標的分子を決定する段階、ならびに

(c) 前記標的分子に結合するかまたはその活性を調節する治療薬を前記データベースから選択する段階により、前記疾病もしくは障害を治療、安定化、または予防する段階を含む方法。

10

【請求項77】

コンピュータで実行される、請求項75記載の方法。

【請求項78】

被験者の疾病または障害の治療、安定化、または予防のための治療法を選択する方法であって、

(a) 治療薬および前記標的分子に結合するかまたはその活性を調節するその能力の記録に関連する、少なくとも10件の該標的分子の記録を含む電子データベースを提供する段階

(b) 被験者において前記疾病または障害に関与する突然変異を有する標的分子を決定する段階、ならびに

20

(c) 前記標的分子に結合しないか、またはその活性を調節しない治療薬を前記データベースから選択する段階を含む方法。

【請求項79】

前記標的分子がタンパク質である、請求項78記載の方法。

【請求項80】

前記標的分子が核酸である、請求項78記載の方法。

【請求項81】

コンピュータで実行される、請求項78記載の方法。

【請求項82】

30

関心対象の化合物が試料中に存在するか否かを決定する方法であって、

(a) 化合物ライブラリから得た2つ以上の化合物の参照質量スペクトルを提供する段階、

(b) 前記ライブラリから得た1つ以上の化合物を含む試料の被験質量スペクトルを提供する段階、および

(c) 参照質量スペクトルのピークが前記被験質量スペクトルに含まれているか否かを決定する段階により、該参照質量スペクトルを生成した化合物が前記試料中に存在するか否かを決定する段階を含む方法。

【請求項83】

前記被験質量スペクトルの全ピークが1つの化合物に割り当てられるまで、前記参照質量スペクトルを順番にまたは同時に解析する、請求項82記載の方法。

40

【請求項84】

段階(c)が、1つ以上の参照質量スペクトルのピークが前記被験質量スペクトルに含まれるか否かを連続して決定する段階を含む、請求項82記載の方法。

【請求項85】

請求項82記載の方法であって、

(i) 前記参照質量スペクトル中の全ピークが前記被験質量スペクトル中に存在すると決定する段階により、該参照質量スペクトルを生成した化合物が前記試料中に存在すると決定する、または

(ii) 該参照質量スペクトルの1つのピークが該被験質量スペクトル中に存在しないと決

50

定する段階により、該参照質量スペクトルを生成した化合物が前記試料中に存在しないと決定する

まで、段階(c)が繰り返される方法。

【請求項86】

段階(a)が、前記ライブラリの各化合物の質量スペクトルを決定する段階を含む、請求項82記載の方法。

【請求項87】

前記参照スペクトル中の少なくとも1つのピークが同位体ピークまたはフラグメントピークである、請求項82記載の方法。

【請求項88】

前記参照スペクトルの少なくとも1つのピークが親ピークである、請求項82記載の方法。

【請求項89】

前記参照質量スペクトルが、前記質量スペクトルを生成した化合物の参照に関連する該質量スペクトルの1つ以上の特性の記録を含むデータベースに含有される、請求項82記載の方法。

【請求項90】

段階(c)がコンピュータで実行される、請求項82記載の方法。

【請求項91】

関心対象の化合物が試料中に存在するか否かを決定する方法であって、

(a) 化合物ライブラリから得た2つ以上の化合物の参照質量スペクトルを提供する段階、
(b) 前記ライブラリから得た1つ以上の化合物を含む試料の被験質量スペクトルを提供する段階、

(c) 前記被験質量スペクトルの1つ以上のピークが参照質量スペクトルに含まれているか否かを決定する段階、および

(d) 参照質量スペクトルの全ピークが前記被験質量スペクトル中に存在するか否かを決定する(ここで該参照質量スペクトルは該被験質量スペクトル中に存在する1つのピークを含む段階(c)由来の参照質量スペクトルである)ことにより、該参照質量スペクトルを生成した化合物が前記試料中に存在するか否かを決定する段階を含む方法。

【請求項92】

段階(d)が、1つ以上の参照質量スペクトルのピークが前記被験質量スペクトルに含まれるか否かを順番に決定する段階を含む、請求項91記載の方法。

【請求項93】

段階(d)が、前記参照質量スペクトルの1つのピークが被験質量スペクトル中に存在するか否かを決定する段階を含む、請求項91記載の方法であって、

(i) 該参照質量スペクトルの全ピークが該被験質量スペクトル中に存在すると決定する段階により、該参照質量スペクトルを生成した化合物が前記試料中に存在すると決定する、または

(ii) 該参照質量スペクトル中の1つのピークが該被験質量スペクトル中に存在しないと決定する段階により、該参照質量スペクトルを生成した化合物が該試料中に存在しないと決定する

まで、該決定が繰り返される方法。

【請求項94】

段階(a)が前記ライブラリの各化合物の質量スペクトルを決定する段階を含む、請求項91記載の方法。

【請求項95】

前記参照スペクトルの少なくとも1つのピークが同位体ピークまたはフラグメントピークである、請求項91記載の方法。

【請求項96】

前記参照スペクトルの少なくとも1つのピークが親ピークである、請求項91記載の方法。

10

20

30

40

50

【請求項 97】

前記参照質量スペクトルが、前記質量スペクトルを生成した化合物の参照に関連する1つ以上の該質量スペクトルの特性の記録を含むデータベースに含有される、請求項91記載の方法。

【請求項 98】

前記特性が、同位体ピークの質量／電荷比、フラグメントピークの質量／電荷比、親ピークの質量／電荷比、およびピーク強度からなる群より選択される、請求項97記載の方法。

【請求項 99】

段階(c)または(d)がコンピュータで実行される、請求項97記載の方法。

【請求項 100】

関心対象の化合物が試料中に存在するか否かを決定するプログラムを内蔵したコンピュータ読み取り可能なメモリであって、

(a) 化合物ライブラリから得た2つ以上の化合物の参照質量スペクトル中、1つ以上のピークの質量／電荷比を含む質量分析データを入力値として受け取るコンピュータコード、
(b) 前記ライブラリから得た1つ以上の化合物を含む試料の被験質量スペクトル中、1つ以上のピークの質量／電荷比を含む質量分析データを入力値として受け取るコンピュータコード、および

(c) 参照質量スペクトル中のピークが前記被験質量スペクトルに含まれるか否かを決定する段階により、該参照質量スペクトルを生成した化合物が前記試料中に存在するか否かを決定するコンピュータコードを含むメモリ。

【請求項 101】

関心対象の化合物が試料中に存在するか否かを決定するプログラムを内蔵したコンピュータ読み取り可能なメモリであって、

(a) 化合物ライブラリから得た2つ以上の化合物の参照質量スペクトル中、1つ以上のピークの質量／電荷比を含む質量分析データを入力値として受け取るコンピュータコード、
(b) 前記ライブラリから得た1つ以上の化合物を含む試料の被験質量スペクトル中、1つ以上のピークの質量／電荷比を含む質量分析データを入力値として受け取るコンピュータコード、

(c) 前記被験質量スペクトルの1つ以上のピークが参照質量スペクトルに含まれるか否かを決定するコンピュータコード、および

(d) 参照質量スペクトル中の全ピークが前記被験質量スペクトル中に存在するか否かを決定する段階により、該参照質量スペクトルを生成した化合物が前記試料中に存在するか否かを決定するコンピュータコードを含むメモリ。

【請求項 102】

関心対象のタンパク質をコードする2つ以上のベクターを作製する方法であって、

(a) 関心対象の第一タンパク質をコードする第一核酸と第一バックボーン核酸とを、それらの反応が生じ得る条件下、ロボット操作装置の第一区画においてロボット操作で接触させる段階により、該第一タンパク質をコードする第一ベクターを作成する段階、および
(b) 関心対象の第二タンパク質をコードする第二核酸と第二バックボーン核酸とを、それらの反応が生じ得る条件下、ロボット操作装置の第二区画内においてロボット操作で接触させる段階により、該第二タンパク質をコードする第二ベクターを作成する段階を含む方法。

【請求項 103】

請求項102記載の方法であって、

(c) 前記第一ベクターと第一細胞とを、該第一細胞内に該第一ベクターが挿入され得る条件下でロボット操作で接触させる段階、および

(d) 前記第二ベクターと第二細胞とを、該第二細胞内に該第二ベクターが挿入され得る条件下でロボット操作で接触させる段階、

10

20

30

40

50

を更に含む、方法。

【請求項104】

前記第一細胞が前記第一タンパク質を発現し、かつ前記第二細胞が前記第二タンパク質を発現する、請求項103記載の方法。

【請求項105】

少なくとも5個のベクターが同時に作製される、請求項102記載の方法。

【請求項106】

タンパク質を精製する方法であって、

(a) ロボット操作装置内の第一培養液中に前記第一タンパク質が分泌される条件下で、第一細胞において該第一タンパク質を発現させる段階、

(b) ロボット操作装置内の第二培養液中に前記第二タンパク質が分泌される条件下で、第二細胞において該第二タンパク質を発現させる段階、

(c) 前記第一培養液を第一クロマトグラフィーカラムに、また前記第二培養液を第二クロマトグラフィーカラムにロボット操作で移す段階、ならびに

(d) 前記第一タンパク質および前記第二タンパク質を精製する段階を含む方法。

【請求項107】

少なくとも5個のタンパク質が同時に精製される、請求項106記載の方法。

【発明の詳細な説明】

【背景技術】

【0001】

発明の背景

1. はじめに

本発明は、標的物質を膨大な数のリガンド中に曝露させ、リガンドー標的物質の対を収集し、そのリガンドを使用してその標的物質の生物学的機能を分析し、状況に応じてリガンドを化学的および/または構造的に同定する方法に関連する。本発明の1つの実施例では、製薬的に適切な標的に結合するリガンドが選択されている。本発明の別の実施例では、リガンドー標的対がゲノムスケールで収集され分析されている。本発明はさらに、表現型の1つの変異に対し少なくとも1つの生物学的アッセイ法で多数の可能なリガンドをスクリーニングし、ヒットしたリガンドを使用しそれに対応する標的分子を同定する方法に関連する。

【0002】

2. 本発明の背景

2.1. 新薬発見への従来法

過去50年間に発見された薬物は一般的に、200~300の標的に基づいており、現在全製薬会社がスクリーニングに使用している有効な標的は総計450個ほど存在している。これらの標的の大体は、典型的に従来の薬物発見方法によって開発されてきており、その方法で、標的は、遺伝子過剰発現、遺伝子ノックアウト、機能ドメインの遺伝子配列相同性検索、X線結晶化学、または特異的な細胞および生物学的アッセイを含む還元主義生物学を使用して確認されている。さらに、今日実施されているような新薬発見では、標的確認、アッセイ開発、ハイスループットスクリーニング、およびリード化合物の生成が連続して実施されている。

【0003】

2.2. ゲノミクス

ヒトゲノム解析の完了により特性の未知な多数の遺伝子の配列が分かり、ヒトゲノム配列の価値を引き出すために正しい標的のみを確認し選択する段階は、製薬企業にとって困難であるが必須なこととなっている。ヒトゲノムの100,000個以上の遺伝子中、最大で10,000個の遺伝子が製薬的に有用な標的になると推定されている。遺伝子のこの膨大な数は、遺伝子確認への還元主義的アプローチを困難にし、その結果薬物発見の主たる障害となっている。

10

20

30

40

50

【0004】

DNA配列データの膨大な蓄積は、この問題の解決を約束する機能ゲノミクスの分野を立ち上げることになった。遺伝子発現のプロフィールはDNAアレイを使用して研究できる (De Risi JL et al., 1997, Science 278 :680)。タンパク質発現プロフィールはタンパク質アレイによって実施可能である (Pawelietz CP et al., 2000, Drug Dev. Research 49:34)。遺伝子機能は、遺伝子を導入・変異させ、表現型に制御変化を誘発させることによって試験できる。これに代わって、遺伝子のアンチセンスまたはリボザイム型が、様々な細胞株や、トランスジェニックマウスまたはノックアウトマウス、線虫 (*C. elegans*)、キンカチョウ、ショウジョウバエまたは酵母などの生物内で発現される可能性がある (Couture LA et al., 1996, Trends in Genetics 12:510; Nadeau JH et al., 1998, Curr. Opin. Genet. Dev. 8, 311)。

【0005】

示差的遺伝子発現は以下のような各種技法によって検出できる。示差スクリーニング (Teder TF et al. 1988 PNAS 85:208)、差し引きハイブリダイゼーション (Hedrick SM et al. 1984, Nature 308:149)、示差ディスプレイ (Liang P and Pardee A 1993 US5262311)、遺伝子マイクロアレイ (Lockhart, D et al., 1996, Nature Biotechnology 14:1675; Schena M et al., 1995, Science 270:467; 2000, Nature Genetics 24:236)、発現差解析法 (representational difference analysis: RDA法) (Hubank M et al., 1994, Nucleic Acids Research 22:5640)、発現配列タグ (EST) の大規模配列決定法、逆転写酵素PCR、遺伝子発現連続分析 (SAGE; Nacht M et al., 1999, Cancer Res. 59:5464) およびレーザー捕獲顕微解剖 (Sgroi DC et al., 1999, Cancer Research 59:5656) などである。マイクロアレイ技法はゲノミクスにおける最新の技術であり、細胞サイクル、生化学経路、酵母のゲノム拡大発現、細胞成長、細胞分化、単一化合物への細胞反応、遺伝病などの研究に利用されてきている (M. Schena, 1998, TIBTECH 16:301)。

【0006】

2.3. 標的タンパク質の同定と特性決定

従来の生化学的技法では、以前未知であった小分子レセプターを、光架橋法、標識リガンド結合、および親和性クロマトグラフィーなどのインビトロ生化学方法によってタンパク質レベルで同定してきた (Jakoby WB et al., 1974, Methods in Enzymology 46:1)。これらの方法はタンパク質の精製を必要とする。レセプター遺伝子をクローン化するためには、ペプチドの配列をそれ以上に決定する必要がある、この配列はそのタンパク質を発現するcDNAをクローン化するのに使用される。小分子は標識化され、その分子標的を決定するために使用される (Kwon HJ et al., 1998, PNAS 95:3356)。代わって、小分子をアガロスマトリックスに固定化し、様々な細胞種および生物の抽出物のスクリーニングに使用できる。例えば、ブルバラノールB (purvalanol B) (サイクリン依存キナーゼ阻害物質として既知) をアガロスマトリックスに固定化し、多様に収集した細胞種と生物の抽出物のスクリーニングに使用し、キナーゼ活性を持つ多数のタンパク質が単離された (Knockaert M et al., 2000, Chem. Biol. 7:411)。一方、トラボキシンは、ヒストン脱アセチル化を阻害し、細胞サイクルを停止させるシクロテトラペプチドである。トラボキシンと共有結合で修飾された親和マトリックス上で、2つの核タンパク質が、分画細胞抽出物からヒストン脱アセチラーゼ活性によって同時に精製された。続いて、これらのタンパク質の配列が決定され、タンパク質をコードするcDNAがcDNAライブラリからクローン化された (Taunton J et al., 1996, Science 272:408)。

【0007】

現時点で、タンパク質-タンパク質間の相互作用を試験する主な系は酵母の2ハイブリッド系である。この方法では、1つのタンパク質がDNA結合ドメインに融合され、別のタンパク質が真核細胞転写因子のDNA活性ドメインに結合され、酵母を成長させるレポーター遺伝子の存在下で発現される。2つの異種タンパク質が2つのドメインを引き合わせると、次いで、相互作用するタンパク質を含んだ酵母が成長によって選択される (Fields S et al., 1989, Nature 340:245)。

10

20

30

40

50

【0008】

酵母の「3ハイブリッド」転写活性化系が、すでに同定された薬物FK506レセプターのコード遺伝子をクローン化するのに使用されている。この3ハイブリッド系は、転写活性化ドメインに融合したcDNAのライブラリに対する、活性リガンドの固着誘導体を表示する (Borchardt A. et al., 1997, Chem. Biol. 4:961; Licitra EJ et al., 1996, PNAS 93:12817)。Licitraらは、ラットの糖質コルチコイドレセプターのホルモン結合ドメインをLex A DNAドメインに融合させ、FK506レセプター(FKBP12) をコードするcDNAを転写活性化ドメインに融合させ、この2つを酵母2ハイブリッド系において発現させた。この酵母細胞はデキサメタゾンとFK506の共有結合ヘテロ二量体を含む培地で平板培養され、細胞は非二量体化FK506によって阻害される可能性のある経路で成長した。この実験を、FK506結合タンパク質をコードするcDNAに代わり、転写活性ドメインに融合したcDNA発現ライブラリで繰り返すと、成長酵母はFK506結合タンパク質をコードするcDNAを含んでいた。しかし、この実験は既知標的と相互作用する化学物質を使用して行っている。Borchardt Aらは、FKBP12-GAL4 DNA結合ドメイン融合、FK506結合タンパク質ラバマイシン関連タンパク質のFRドメイン、およびラバマイシンなどの存在下で、酵母細胞にヒスチジン不在下で細胞を成長させてHIS3レポーター遺伝子3個を転写させている。

10

【0009】

発現クローニングは、少数プールのタンパク質内で標的を試験するために使用され得る (King RW et. al., 1997, Science 277:973)。ペプチド (Kieffer et. al., 1992, PNAS 89:12048)、ヌクレオチド誘導体 (Haushalter KA et. al., 1999, Curr. Biol. 9:174)、および薬物-ウシ血清アルブミン (薬物-BSA) 結合体 (Tanaka et. al., 1999, Mol. Pharmacol. 55:356) が発現クローニングに使用されてきた。

20

【0010】

標的コードDNAとリガンド結合に密接に関与する別な有用な技法は、ファージディスプレイである。ファージディスプレイは、モノクローナル抗体分野で主に利用されてきたが、ペプチドまたはタンパク質ライブラリがウィルス表面で作成され、活性がスクリーニングされる (Smith GP, 1985, Science 228:1315)。ファージは固相に接続された標的用により分けられる (Parmley SF et al., 1988, Gene 73:305)。ファージディスプレイの利点の1つは、cDNAがファージ内に存在しているので、別のクローニング手順を必要としないことである。Dyaxはファージディスプレイ親和カラムを使用して、小分子ではなく、マクロ分子を単離した (US97/04425)。

30

【0011】

最近、Scheらは親和性プローブとして天然物質であるFK506を使用し、T7 cDNAファージディスプレイライブラリからFKBP12をクローン化した。彼らは、ビオチン化したFK506を支持する親和性マトリックスを使用して、ヒト脳cDNAから調製したファージライブラリをスクリーニングした。2ラウンドの親和性選択後、残存するファージ粒子は、完全な長さのFKBP12に対応する通常の450 bp 挿入部を共有していた。

【0012】

ファージディスプレイの代わりとして、プラスミドディスプレイ (Cull et al., 1992, PNAS 89:1865; Schatz PJ et al., 1996, Methods Enzymol 267:171)、ポリソームディスプレイ (Mattheakis LC et al., 1996, PNAS 91:9022; Mattheakis LC, 1996, Methods Enzymol 267:195)、タンパク質タギング (Whitehorn EA et al., 1995, Biotechnology 13:1215)、リボソームディスプレイ (Hanes J et al., 1998, PNAS 95:14130)、および細菌および真核生物の細胞表面ディスプレイ (Georgiou G et al., 1997, Nat. Biotechnol 15:29; Chesnut J et. al., 1996, J. Imm Methods 193:17) などがある。ペプチドまたはタンパク質も、プロマイシンをコードするmRNAに、プロマイシンを介して化学的に連結することができる (Roberts R et al., 1997, PNAS 94: 12297)。

40

【0013】

2.4. 化学遺伝学

化学遺伝学は化学物質を使用して遺伝子機能を確認し、遺伝子発現や遺伝子機能に制御的

50

変化を起こすための新たな強力と考えられるアプローチである。しかし、今日まで、その化学遺伝学的アプローチも、その対象薬物がすでに市場に出回っている既知標的を用いたハイスループット細胞に基づく従来のスクリーニングアッセイによって、これらの既知標的にヒットするより多くの物質を見出すという、伝統的な薬物発見プロセスから飛躍的な進歩は見られていない。化学遺伝学の現状は、Haggarty SJらによる研究(2000, Chem Biol 7:275)によって説明されるが、その研究では139種類の化合物が、細胞に基づくアッセイにおいて、有糸分裂阻害に対するChembridge Diversetライブラリのハイスループットスクリーニングから同定され、続いてインビトロチューブリン重合アッセイ法によって分析された。139個の化合物のうち52個はコルヒチンと同じメカニズムでチューブリンを不安定にするアンタゴニストであった。1つの化合物はタキソールと同様なメカニズムによって、チューブリンを安定化させるアゴニストであることが実証された。86個の化合物には作用はなく、非チューブリン標的を通じ、有糸分裂をモジュールする傾向であった。染色体および細胞骨格に対する目で見える作用に基づいて、非チューブリン標的をターゲットする化合物に対し、7個がチューブリンの弱いアンタゴニストであると考えられており、1個(モノステロール)はキネシン関連タンパク質Eg5を阻害することが認められた(Mayer et. al., 1999, Science 286:971)。Haggarty SJらの実験では、20~50 μ Mのリガンド濃度でアッセイが実施されたことから、低親和性リガンドが選択されている。しかし、標的機能を決定する段階では、低親和性リガンドの価値は限定される。

【0014】

Rosania GRらはチューブリンに結合し、筋肉細胞の可逆的核分裂や増殖を誘発させる細胞形態学スクリーニング法によって、新規な小分子、ミロセベリン(myoseverin)を同定した。この最新の発明とは異なり、Schulzは標準の機能ゲノミクスであるDNAアレイ法の力をかりて、メカニズムを解明している(Rosania GR et. al., 2000, Nat Biotechnol 18:304)。1889年にコルヒチンが有糸核分裂に影響を与えることが認められて以来、化学物質が機能解明に使用されてきている(Eigsti O, 1949, Science 110:692)。しかし現在、実際には、既知標的あるいは特定の表現型を生ずる未同定標的に結合するリガンドを同定するだけに留まっている。

【0015】

未知遺伝子の機能特性を知るために過去に行われた努力の結果は、オーファン(みなしご)レセプター解析によって裏付けられている。オーファンレセプターは、以前に同定されたレセプターとDNA配列に類似性を持つ遺伝子によってコードされる。これに基づいて、それらの配列は自然の生理的役割やリガンドが未知なレセプターのスーパーファミリー内に挿入される。現時点での最新技術は、遺伝学技術を使用するか、あるいはファミリーの他のメンバーに結合することが知られている薬物またはタンパク質リガンドを使用して、それらの機能を決定する段階である(Werme M et. al., 2000, Brain Res 863:112; Bordji K. et. al., 2000, J. Biol. Chem. 275:12243; Yang C., 1999, Cancer Res. 59:4519; Chiou L, 1999, Br. J. Pharmacol 128:103; Williams C, 2000, Curr. Opinion in Biotechnology 11:42)。

【0016】

2.5. 標的化学物質の特性決定

一旦標的物質が確認されると、生物学的アッセイとメカニズムに基づくアッセイの2つの主なスクリーニング法が適用される(Gordon et. al., 1994, J. Med. Chem. 37:1386)。生物学的アッセイでは、生存度または代謝によってスクリーニングされた化合物の細胞1個に対する作用を測定する。例えば、ペニシリンは細菌培養中、成長が抑制されたことで発見された。メカニズムに基づくアッセイには、酵素活性への作用を測定する生化学的アッセイ、標的およびレポーターシステム(例えば、ルシフェラーゼまたは β -ガラクトシダーゼ)が1つの細胞に導入された細胞ベースのアッセイ(Monks A et. al., 1997, Anticancer Drug Des. 12: 533)、または結合アッセイなどがある。結合アッセイは、ウェル、ピーズ(Bosworth N et al., 1989, Nature 1989, 341:167; Meldal M, 1994, PNAS 91, 3314)、またはチップ(Sunberg S, 2000, Curr. Opin. In Biotechnol 11:47)に固

10

20

30

40

50

定された標的、あるいは固定化抗体によって捕獲された標的を用いて実施され、通常、結合したリガンドは熱量計により、または蛍光を測定することにより検出される (Sunberg S, 2000, Curr. Opin. In Biotechnology 11:47)。

【0017】

新たな結合アッセイの中には、既知機能を持つ標的に結合する分子をキャピラリー電気泳動によって分離する方法もある (US 5783397; US99/15458)。他の新規アッセイ法では、ライブラリを質量スペクトルによって分子量によりコード化し逆重量積分している (Carell T et al., 1995, Chem Biol. 2: 171; Fang AS et. al., 1998, Comb Chem High Throughput Screen 1:23; US 99/23837; US99/00024)。組み合わせライブラリの純度を測定し、血漿試料中の代謝物を分析するために、質量スペクトルと共にHPLCも使用されている (Korfmacher WA et al., 1999, Rapid Commun Mass Spectrom 13:1991; Zeng L et al., 1998, Comb Chem High Throughput Screen 1:101; Nedved ML et al., 1996, Anal Chem 68: 4228; Zimmer D et al., 1999, J. Chromatogr A 854:23; Aubagnac JL, Comb Chem High Throughput Screen 2:289)。

【発明の開示】

【0018】

3. 本発明の要旨

本発明は、機能未知な標的物質を使用し、続いてアッセイで使用する化学物質ライブラリから小分子を選択し、標的の機能を決定する段階に関連する。本発明によれば、化学ライブラリのメンバーを生化学的結合アッセイによりタンパク質と混合し、続いて結合するメンバーを（順番にまたは同時に）インビトロまたはインビボで生物学的アッセイを行い、生物学的あるいは病理学的条件下で、測定可能な表現型の変化から遺伝子機能を決定する。

【0019】

また、本発明では、生物学的アッセイで表現型の変化を誘発する化学物質を使用して、標的物質の識別を決定する。本発明は、少なくとも1つの生物学的アッセイで多数の可能性のあるリガンドをスクリーニングし、1つの生物学的アッセイで表現型の変化を生ずるリガンドを選択し、さらにそのリガンドを使用して標的候補物質をスクリーニングし、変化した表現型の原因である特定の標的物質を同定する方法を提供している。

【0020】

本発明を使用して、遺伝子機能を確定し、同時に薬物標的を確認して薬物リード化合物を得ることにより、薬物発見方法を合理化することができる。構造活性相関の情報は、標的物質に結合するが表現型アッセイで異なる活性を示す、多数の多様な構造を有するヒット化合物を、同時に比較することにより得られ、その情報はリード化合物を迅速に最適化するのに使用できる。本発明によって、ゲノミクスから得た膨大な数の遺伝子が系統的に分類され、特定の疾患に対する有用な薬物標的が確認され、選択され得る。

【0021】

本発明は現在の技術とは異なっている、なぜなら現在の技術では既知標的についてスクリーニングし、一方本発明は標的の正体または機能に関して従前の知識を必要としないからである。さらに、本発明は、ライブラリ構築において、あらかじめ決定された特定質量のサブユニットによる制約は全く受けない。本発明によれば、組み合わせ法または単独法によって生成された、実質上すべてのリガンドライブラリを使用出来ると思われる。制限のない例として、化学物質、ペプチド、天然物質、天然様類似物質、糖質または抗体ライブラリが挙げられる。HIV TAT, HSV VP22またはタンパク質形質導入ドメインを含んだアンテナペディア (Antennapedia: 黄色ショウジョウバエ変異遺伝子群) ペプチドの1つの配列を利用してペプチドおよびタンパク質が細胞膜を交差するようにできる (Swartz SR et al., 2000, Trends in Cell Biology 10:290)。ライブラリは恐らくリガンドのプールから構成されるか、または個々にスクリーニングされた単一リガンドの収集されたものである。

【0022】

従って、1つの局面として、本発明は標的分子に結合するリガンド候補を選択する方法を特徴としている。その方法は、標的分子と1つ以上の候補リガンドとの複合体が形成できる条件下で、候補リガンドのライブラリを持つ標的分子を含むインビトロ試料との接触に関する。この複合体が単離され、その中から1つ以上の候補リガンドが回収される。さらに、回収された候補リガンドの1つ以上が同定される。

【0023】

上記局面に対する様々な態様において、標的分子は未知の生物学的機能を持つ分子であるかまたは薬物標的として以前に確認されていない分子である。他の実施例は、ライブラリに少なくとも2つの異なる骨格 (scaffolds) を、または少なくとも11種の異なる化合物を含んでいる。他の実施例では、複合体がサイズ除外法または2層クロマトグラフィーによって単離されている (すなわち、内部界面逆相 (ISRP)、GFF、またはGFFII樹脂を使用したクロマトグラフィー)。他の実施例では、MS、IR、FTIR、NMR、および/またはUV分析法を使用して、回収した候補リガンドを同定している。他の実施例では、本発明の方法として、回収候補リガンドの質量スペクトルにおいて、親ピーク、フラグメントピーク、および/または同位体ピークの質量/電荷比を決定する方法も含んでいる。1つの実施例には、その標的分子に結合することが既知な競合リガンドを持つ試料との接触方法も含んでいる。この競合リガンドは標的分子と結合する低親和性の候補リガンドの数を減少させ、より高い親和性の候補リガンドが選択されるようにできる。

【0024】

1つの局面として、本発明は、標的分子に結合するリガンド候補を選択する別な方法を提供とする。この方法では、第一標的分子と1つ以上の候補リガンドとの複合体が形成され、また第二標的分子と1つ以上の候補リガンドとの複合体が形成される条件下で、第一標的分子と第二標的分子をリガンド候補ライブラリと共に含んだインビトロ試料との接触に関する。候補リガンドと結合した第一標的分子を含む複合体、および候補リガンドと結合した第二標的分子を含む複合体が単離される。第一複合体および/または第二複合体から1つ以上の候補リガンドが回収され同定される。1つの実施例には、第一標的分子または第二標的分子に結合することが既知な、競合リガンドと試料との接触方法も含んでいる。

【0025】

さらに、本発明は、天然にまたは人工的に存在するタンパク質、核酸、炭水化物あるいは他の有機分子など、標的分子の生物学的に機能を決定する様々な方法を提供する。この方法は、特定の疾病状態または特定の生物学的刺激物質 (例えば、TNF α) の存在下で上方制御または下方制御される遺伝子またはタンパク質のような、対象とする遺伝子またはタンパク質の機能を決定するために使用され得る。この方法はまた、疾病状態の治療において、治療に有効な化合物を同定するために使用され得る。

【0026】

このような1つの局面において、本発明は標的分子の生物学的機能を決定する方法を提供する。この方法は、1つ以上の候補リガンドが標的分子と複合体を形成できる条件下で、標的分子と候補リガンドのライブラリを含むインビトロ試料との接触に関する。標的分子と結合する候補リガンドが選択される。生物学的アッセイによって、選択した候補リガンドの作用が測定され、続いて標的分子の生物学的機能が決定される。様々な実施例において、標的分子とは、生物学的機能が未知の分子であるか、または薬物標的として以前に確認されていない分子である。他の実施例において、標的分子は、疾病状態、生理学的刺激物質の存在 (例えば、TNFのようなサイトカイン) 下、または特定の細胞または生物学的過程において、上方制御または下方制御される。特定の実施例において、標的分子は、脈管形成、分化、増殖、またはインスリン分泌中に上方制御または下方制御される。1つの実施例において、選択された候補リガンドは、MS、IR、FTIR、NMR、UVまたは他の適切な方法によって同定される。特定の実施例において、選択された候補リガンドにより生物学的アッセイ中に標的分子が増加する。例えば、候補リガンドは、標的分子の作用 (酵素活性のような) を活性化し、標的分子の生成を促進し、標的分子の安定性を高め、標的分子の位置を変化させ、または標的分子と別の分子との会合を促進させることもある。他

10

20

30

40

50

の実施例において、選択された候補リガンドは生物学的アッセイで測定すると標的分子の活性を低下させる。例えば、候補リガンドは、標的分子の活性を阻害し、標的分子の生成を阻害し、標的分子の安定性を低下させ、標的分子の位置を変化させ、または標的分子と別の分子との会合を抑制させることもある。代表的な生物学的アッセイには、非形質導入細胞株、細胞、組織、またはその標的が以前に知られていない他の生物学的システムによるスループットスクリーニングを含む。他の実施例において、生物学的アッセイには、疾病または異常を持つ生物、または生理学的刺激物質の有無にかかわらず特異的な細胞または生体過程が起こっている生物の1組織に対する選択した候補リガンドの作用を決定し、その結果、標的分子の生物学的機能を決定する段階が含まれる。1つの実施例における組織は、ヒト組織のような哺乳類組織である。

10

【0027】

同じ標的分子に結合する2つのリガンドを架橋する方法も提供されている。これらの方法により、1つ以上の標的物質表面上で、2つのリガンドの反応が促進または触媒される。これらの方法をリガンドライブラリのスクリーニングに用い、どのリガンドが標的分子と結合するか、またリガンドと組んだどの架橋生成物が最も高い親和力で標的分子と結合するかを決定する段階もある。架橋生成物を、治療法開発においてリード化合物として使用するか、または標的分子の活性部位を特徴づけるために使用することもある。異なった標的分子に結合する2つのリガンドを架橋させるのに、関連方法を使用することもある。これらの方法を、どの標的分子が対象標的分子と相互作用するかを決定し、その結果、どの分子が対象標的分子と同じ経路に存在するかを決定するために使用することもある。

20

【0028】

別の局面において、本発明は、関心対象の標的分子と結合する2つのリガンドを反応させる方法の特徴としている。この方法は、標的分子が第一リガンドと第二リガンド両方に結合し、さらに最初の架橋剤が第二リガンドと共有結合できる条件下で、第一リガンド（例えば、最初の架橋剤を有する第一リガンド）と第二リガンドを伴う標的分子を含む細胞またはインビトロ試料に接触し、その結果、第一リガンドと第二リガンドを含む架橋生成物の形成に関する。実施例によっては、標的分子は未知の二次および三次構造を有する分子である。他の実施例では、第一リガンドまたは第二リガンドの標的分子との結合部位の位置または三次構造は知られていない。特定の実施例では、架橋生成物の標的分子に対する親和性は、第一リガンドや第二リガンドよりも高い。別の実施例で、架橋生成物は、新薬発見または開発、リード化合物の最適化、または農業用または環境用の物質の開発に使用される。さらに別の実施例で、標的分子は、第一と第二リガンド間の反応を促進または触媒する。別の実施例で、第一リガンドは、標的分子と接触する前に架橋剤の反応を受ける。さらに別の実施例で、第一リガンド、第二リガンドおよび架橋剤は、標的分子の有無にかかわらず反応を受ける。

30

【0029】

別の局面において、本発明は、異なる標的分子と結合する2つのリガンドを反応させる方法の特徴とする。この方法は、第一リガンド（例えば、最初の架橋剤を有する最初のリガンド）および第二リガンドを伴う、第一標的分子と第二標的分子を含む細胞またはインビトロ試料との接触を含む。接触は、(i)第一標的分子を第一リガンドに結合させる、(ii)第二標的分子を第二リガンドに結合させる、(iii)第一架橋剤を第二リガンドと共有結合させるという条件下で、実施され、その結果、第一リガンドと第二リガンドを含む架橋生成物が形成される。1つの実施例において、第一標的分子と第一リガンドとの結合部位の位置または三次構造、および/または第二標的分子と第二リガンドとの結合部位の位置または三次構造は知られていない。1つの実施例において、架橋生成物が形成することは、第一標的分子（例えばタンパク質）と第二標的分子（例えばタンパク質）が、インビボで相互作用を行うか、あるいは同じ生体経路の一部であることを示している。別の実施例で、架橋生成物は、新薬発見または開発、リード化合物の最適化、または農業用または環境用の物質の開発に使用される。さらに別の実施例で、1つまたは両方の標的分子は、第一と第二リガンド間の反応を促進するか、または触媒する。別の実施例で、第一リガンドは

40

50

、標的分子と接触する前に架橋剤の反応を受ける。さらに別の実施例で、第一リガンド、第二リガンドおよび架橋剤は、標的分子の有無にかかわらず反応を受ける。

【0030】

別の局面において、本発明は、第一タンパク質と結合する第二タンパク質を単離する方法を提供する。この方法は、第一架橋剤を持つ第一リガンドおよび第二リガンドを伴う第一タンパク質と第二タンパク質を含む、細胞またはインビトロ試料との接触に関する。その接触は、(i)第一タンパク質を第一リガンドに結合させる、(ii)第二タンパク質を第二リガンドに結合させる、(iii)第一架橋剤を第二リガンドと共有結合させるという条件下で、実施され、その結果、第一リガンドと第二リガンドを含む架橋生成物が形成され、さらに架橋生成物、第一タンパク質および第二タンパク質を含む複合体が形成される。この複合体は単離され、複体内の、あるいは複合体から回収された第一タンパク質および/または第二タンパク質が同定される。1つの実施例において、第一および/または第二タンパク質は検出可能な基を含んでいる。別な実施例において、第二リガンドは架橋剤を含む。1つの実施例において、架橋生成物が形成することは、第一タンパク質と第二タンパク質が、インビボで相互作用を行うか、あるいは同じ生体経路の一部であることを示している。別の実施例で、架橋生成物は、新薬発見または開発、リード化合物の最適化、または農業用または環境用の物質の開発に使用される。

【0031】

本発明はまた、関心対象の化合物と結合する標的分子を選択する多数の方法を提供する。例えば、その化合物は、疾病状態を促進または阻害するように見える1つの分子であるかもしれない。選択された標的分子は、例えば、疾病の研究、疾病に関与する他の分子の同定、また標的分子に結合するか、またはその活性を調節する治療法の同定、あるいは疾病経路の別のメンバーの同定に、使用されることもある。

【0032】

別の局面で、本発明は、関心対象の小分子と結合する標的分子候補を選択する方法を提供する。その方法は、関心対象の小分子と1つ以上の標的分子候補との複合体が形成できる条件下で、標的分子候補のライブラリを持つ小分子を含むインビトロ試料との接触に関する。複合体は単離され、複合体から1つ以上の標的分子候補が回収され、その結果、関心対象の小分子と結合する1つ以上の標的分子候補が選択される。様々な実施例で、標的分子候補ライブラリは、遺伝子組み換え的に生成されるか、または細胞、組織あるいは生物の抽出物から得られる。標的分子候補のライブラリは、関心対象の小分子との接触に先立ち、他の構成物質から精製されないか、部分的に精製されるか、または完全に精製され得る。様々な実施例で、標的分子はフェージ表面で発現されるか、または発現されない。1つの実施例で、標的分子候補ライブラリを持つ小分子との接触に先立ち、対象とするその小分子は、生物学的アッセイにおける作用に基づいて小分子ライブラリから選択される。1つの実施例で、この方法は、選択された標的タンパク質の同定も含む。特定の実施例で、関心対象の小分子は、アミノ酸以外の部分を持つか、または5000、4000、3000、2000、1000、750、500、または250ダルトン未満の分子量を持つ。

【0033】

別の局面で、本発明は、関心対象の小分子と結合する、標的タンパク質を選択する方法を提供する。この方法は、細胞集団内で、表面タンパク質に共有結合した標的タンパク質を含むタンパク質融合の発現、すなわち、細胞表面上でタンパク質融合がディスプレイされる条件下で行われる発現を含む。細胞は関心対象の小分子と接触し、関心対象の小分子と結合する細胞が選択され、その結果、その関心対象の小分子と結合する標的タンパク質が選択される。代表的な細胞は、哺乳類、細菌、酵母および昆虫の細胞である。1つの実施例で、この方法は、選択された標的タンパク質の同定も含む。特定の実施例で、関心対象の小分子は、アミノ酸以外の部分を持つか、または5000、4000、3000、2000、1000、750、500、または250未満の分子量を持つ。

【0034】

別の局面において、本発明は、関心対象の小分子と結合する標的タンパク質を選択する別

10

20

20

30

30

40

40

50

50

物の作用を測定し、その結果、標的分子の生物学的機能を決定する段階も含む。さらに他の実施例で、その細胞またはインビトロ試料は、少なくとも2、5、10、20、30、50、100個、またはそれ以上の標的分子を含み、各標的分子に対し、標的分子と結合あるいはその活性を調節する候補化合物が選択される。

【0039】

別の局面で、本発明は標的分子と結合するか、またはその活性を調節する候補化合物を選択する方法を特徴とする。この方法は、1つ以上の候補化合物を第一標的分子と結合させるか、またはその活性を調節させる条件下で、また1つ以上の候補化合物を第二標的分子と結合させるか、またはその活性を調節させる条件下で、候補化合物のライブラリを持つ第一および第二標的分子を含む、細胞またはインビトロ試料への接触に関する。第一標的分子と結合するかまたはその活性を調節する候補化合物が選択され、さらに第二標的分子と結合するかまたはその活性を調節する候補化合物が選択される。1つの実施例で、1つ以上の選択された候補化合物が同定される。他の実施例で、この方法は、選択した1つ以上の候補化合物の作用を生物学的アッセイにより測定し、その結果、標的分子の生物学的機能を決定する段階も含む。さらに他の実施例で、その細胞またはインビトロ試料は、少なくとも5、10、20、30、50、100個、またはそれ以上の標的分子を含み、各標的分子に対し、標的分子と結合あるいはその活性を調節する候補化合物が選択される。

【0040】

本発明はまた多様なデータベースを特徴とする。これらのデータベースは、本発明のいずれかの方法によって得られる情報を保存するのに有用である。これらのデータベースは、治療法の開発、および特定な患者または特定なタイプの患者に対する好ましい治療法を選択する際に使用されることもある。これらのデータベースの他の利用法が、多数本書に記載されている。

【0041】

上述の1つの局面において、本発明は、標的分子と結合するかあるいはその活性を調節するリガンドおよびそのリガンド能力の記録に関連する標的分子の記録が、少なくとも 10^1 、 10^2 、 10^3 、 10^4 、 10^5 、 10^6 、 10^7 、 10^8 、または 10^9 件含む電子データベースを特徴とする。

関連する局面として、本発明は、生物学的機能が未知で薬物標的および/または標的分子として以前に確認されていない、標的分子の膨大な記録を含む電子データベースを提供する。それらの記録は、標的分子と結合するか、またはその活性を調節するリガンドとその能力に関する記録に関連している。別の関連する局面において、本発明は、ドメインと結合するリガンドとその能力についての記録に関連する、標的分子ドメインの記録が、少なくとも 10^1 、 10^2 、 10^3 、 10^4 、 10^5 、 10^6 、 10^7 、 10^8 、または 10^9 件含む電子データベースを特徴とする。「ドメイン」とは、同じタイプの反応を触媒するか、または同じタイプの分子と結合する1つ以上のタンパク質中に見られるドメインを意味し、このドメインは、DNAあるいはアミノ酸配列分析、X線結晶構造解析、または生物学的アッセイに基づき、異種タンパク質構造モチーフあるいは機能ファミリーとして同定される。例えば、データベースは、キナーゼドメイン（すなわち、1つ以上のキナーゼと結合する能力）またはホスファターゼドメイン（すなわち、1つ以上のホスファターゼと結合する能力）に結合するリガンドおよびその能力についての記録を含むことがある。このデータベースは、例えば、タンパク質または他の標的分子の結合部位の特性決定に、また特定な結合部位または特定な化合物族へのリガンドの選択性を決定する際に使用されることもある。

【0042】

上記データベースの様々な態様において、データベースは、細菌、酵母または哺乳類のような生物のプロテオーム中の、タンパク質あるいはタンパク質ドメインの少なくとも0.5、1、5、10、20、30、40、50、60、70、80、90、または100%の記録を含んでいる。特定な実施例で、データベースは、ヒトプロテオーム中のタンパク質あるいはタンパク質ドメインの少なくとも0.5、1、5、10、20、30、40、50、60、70、80、90、または100%の記録を含んでいる。さらに別の実施例で、データベースは、1つの生物のゲノム中の転写読取枠（open reading frames : ORF）の少なくとも0.5、1、5、10、20、30、40、50、60、70、

80、90、または100%に対して、ORFによって発現された少なくとも1つのタンパク質の記録を含む。

【0043】

別の局面で、本発明は、本発明のデータベースを含み、さらに(i) その記録がコンピュータに保存されている標的分子に結合するか、あるいはその分子の活性を調節する1つ以上のリガンドを表示できる、または、(ii) その記録がコンピュータに保存されているリガンドに結合するか、あるいはそのリガンドによって調節される活性を有する1つ以上の標的分子を表示できる、ユーザーインタフェースを含むコンピュータを特徴とする。代表的なデータベースは、以前に確認されていない標的分子、あるいは生物学的機能が未知な標的分子のような標的について少なくとも10件の記録を含んでいる。

10

【0044】

別の局面で、本発明は、化合物によって影響を受ける1つ以上の生物学的アッセイにおいて、1つの表現型の記録に関連して、少なくとも 10^2 、 10^3 、 5×10^3 、 10^4 、 10^5 、 10^6 、 10^7 、 10^8 、または 10^9 件の化合物の記録を含む電子データベースを提供する。生物学的アッセイは、化合物と結合するタンパク質をコードする核酸の外因性コピーを含まないか、あるいは外因性リポーター遺伝子を含まない、細胞またはインビトロ試料に関する。

【0045】

別な局面で、本発明は、上述の局面におけるデータベースを含み、さらに(i) その記録がコンピュータに保存されている化合物に対する1つ以上の生物学的アッセイ中で1つ以上の表現型を表示できるか、または、(ii) その記録がコンピュータに保存されている表現型に作用する1つ以上の化合物を表示できる、ユーザーインタフェースを含むコンピュータを特徴とする。

20

【0046】

別の局面で、本発明は、標的分子の発現プロフィールまたは活性の記録に関連して、少なくとも10件の標的分子記録を含む電子データベースを提供する。別の局面として、本発明は、標的分子の発現プロフィールまたは活性の記録に関連する、機能が未知で、薬物標的および/または標的分子として以前に確認されていない標的分子の膨大な記録を含む、電子データベースを特徴とする。いずれのデータベースの様々な実施例において0.5、1、5、10、20、30、40、50、60、70、80、90または100%の生物プロテオームのタンパク質の記録、または、少なくとも 10^2 、 10^3 、 5×10^3 、 10^4 、 10^5 、 10^6 、 10^7 、 10^8 、または 10^9 件の標的分子の記録を含む。他の実施例では、データベースは、生物プロテオーム（例えばヒトプロテオーム）中の少なくとも0.5、1、5、10、20、30、40、50、60、70、80、90、または100%のタンパク質の記録を含む。さらに別の実施例で、データベースは、1つの生物のゲノム中の転写読取枠（open reading frames : ORF）の少なくとも0.5、1、5、10、20、30、40、50、60、70、80、90、または100%に対して、ORFによって発現された少なくとも1つのタンパク質の記録を含む。

30

【0047】

さらに別の局面で、本発明は、本発明のデータベースを含み、さらに(i) その記録がコンピュータに保存されている標的分子の、1つ以上の発現プロフィールまたは活性を表示できるか、または、(ii) その記録がコンピュータに保存されている発現プロフィールまたは活性を有する、1つ以上の標的分子を表示できる、ユーザーインタフェースを含むコンピュータを特徴とする。様々な実施例において、データベースは、以前に薬物標的として確認されていない標的分子、あるいは生物学的機能が未知な標的分子のような標的分子の少なくとも10件の記録を含んでいる。

40

【0048】

データベースまたはコンピュータは共に、以下のいずれかの方法で使用できる。これらのデータベースの代表的な使用法には、活性部位／タンパク質の共有する化学的骨格構造と種類の収集、結合特徴と重なりなど結合特性の広範囲なインデックス作成、標的分子の骨格の特異性の確認、化合物の潜在的毒性の確認、特定な生物学または病理学を探索する化合物の選択、特定な化合物の作用の原因となる標的分子の選択、薬理ゲノミクスに基づく

50

治療法の選択、および薬物最適化のリード化合物として必要な骨格を選択する段階などを含む。

【0049】

そういった1つの局面において、本発明は、関心対象の表現型に關与する標的分子を同定する方法を特徴とする。この方法は、リガンドおよび表現型を引き起こすかまたはその一因となるリガンド能力の記録に關連する、生物学的アッセイにおける表現型の膨大な記録を含んだ電子データベースの使用に關する。関心対象の表現型の選択が受け入れられ、その表現型を導く一因となる1つ以上のリガンドが同定される。リガンドに結合するか、またはリガンドによって調節される活性を有する標的分子の記録に關連する、膨大なリガンドの記録を含む電子データベースを使用して、関心対象の表現型を導く一因となるリガンドに結合するか、またはそのリガンドによって調節される1つ以上の標的分子を同定し、その結果、関心対象の表現型に關与する1つ以上の標的分子を同定する。1つの実施例で、関心対象の表現型は疾病状態に關連しており、標的分子がその疾病状態を促進するかまたは阻害するかが決定される。1つの実施例で、この方法はコンピュータで実行される。

10

【0050】

さらに別の局面において、本発明は、関心対象の標的分子に会合している表現型を同定する方法を特徴とする。この方法は、標的分子に結合するか、またはその活性を調節するリガンドおよびリガンド能力の記録に關連する、膨大な標的分子の記録を含み、さらに関心対象の標的分子の選択を受け入れる電子データベースを提供することに關する。標的分子に結合するかその活性を調節する1つ以上のリガンドが同定される。生物学的アッセイ中のリガンドによりもたらされる表現型の記録に關連する、膨大なリガンドの記録を含む電子データベースが提供され、そのデータベースを使用して、その生物学的アッセイ中の1つ以上のリガンドによりもたらされる表現型を同定し、その結果、関心対象の標的分子に關与する1つ以上の表現型を同定する。1つの実施例で、この方法はコンピュータで実行される。

20

【0051】

さらに別の局面において、本発明は、結合するか、または関心対象の標的分子の活性を調節するリガンドを同定する方法を特徴とする。この方法は、標的分子に結合するか、またはその活性を調節するリガンドおよびリガンド能力の記録に關連する、少なくとも10件の標的分子の記録を含み、さらに関心対象の標的分子の選択を受け入れる電子データベースを提供することに關する。標的分子に結合するかその活性を調節する1つ以上のリガンドが同定される。様々な実施例において、本方法は、結合するか、または関心対象の標的分子活性を調節する2つ以上の化学構造を比較し、その結果、関心対象の標的分子の結合または調節を促進するリガンドが持つ官能基を同定する段階を含む。他の実施例では、本方法は、関心対象の標的分子と結合するか、またはその活性を調節する2つ以上のリガンドの化学構造を比較し、その結果、リガンドの収集中に1つ以上の官能基または骨格の頻度を決定する段階も含む。他の実施例では、1つ以上の化合物が、薬物の発見または開発、あるいはリード化合物の最適化において使用する、2つ以上のリガンドに存在する1つ以上の官能基を持っている。1つの実施例で、この方法はコンピュータで実行されている。

30

【0052】

さらに別な局面において、本発明は、結合するか、または関心対象のリガンドによって調節される活性を有する標的分子を同定する方法を特徴とする。この方法は、結合するか、またはそのリガンドによって調節される活性を有する標的分子の記録に關連する、少なくとも10件のリガンドの記録を含み、さらに関心対象のリガンドの選択を受け入れる電子データベースを提供することに關する。結合するか、または関心対象のリガンドによって調節される活性を有する1つ以上の標的分子が同定される。様々な実施例において、本方法は、関心対象のリガンドと結合する2つ以上の標的分子の化学構造を比較し、その結果、関心対象のリガンドの結合を促進あるいは、その結合の一因となる標的分子内の官能基あるいはドメインを同定する段階を含む。

40

【0053】

50

さらに別の局面では、本発明は、関心対象のリガンドの選択性を決定する方法を特徴としている。この方法は、標的分子に結合するか、またはその活性を調節するリガンドおよびリガンド能力の記録に関連する、少なくとも10件の標的分子の記録を含み、さらに関心対象のリガンドの選択を受け入れる電子データベースを提供することに関する。結合するかまたはリガンドによって調節される、データベース中の標的分子の数が決定され、その結果、関心対象のリガンドの選択性を確認する。様々な実施例において、リガンドは、疾病状態、有害な副作用、または毒性に関与している標的分子の活性を増大させるが、そういったリガンドは、薬物の発見または開発、リード化合物の最適化、または農業用または環境用の物質の開発から除外される。他の実施例において、リガンドは、疾病状態、有害な副作用、または毒性に関与している標的分子の活性を抑制させるが、そういったリガンドは、薬物の発見または開発、リード化合物の最適化、または農業用または環境用の物質の開発向けに選択される。1つの実施例で、この方法はコンピュータで実行されている。

10

【0054】

さらに別な局面では、本発明は、被験者の疾病もしくは障害の治療、安定化、または予防のための治療法を選択する方法を提供する。この方法は、治療薬および標的分子に結合するかまたはその活性を調節する能力の記録に関連する、少なくとも10件の標的分子を含む電子データベースを提供し、さらに疾病または障害に関与する突然変異を有する被験者において標的分子を決定する段階を含む。標的分子と結合するかまたはその活性を調節する治療薬がデータベースから選択され、その結果、治療薬により疾病または障害が治療、安定化または予防される。他の実施例で、突然変異を有する被験者または被験者群が、その治療法の臨床試験のために選択されるか、またはその臨床試験の特定サブグループに分類される。特定な実施例で、標的分子は、タンパク質または核酸である。1つの実施例で、この方法はコンピュータで実行されている。

20

【0055】

さらに別な局面では、本発明は、患者の疾病もしくは障害の治療、安定化、または予防のための治療法を選択する別法を特徴とする。この方法は、治療薬および標的分子に結合するかまたはその活性を調節する能力の記録に関連する、少なくとも10件の標的分子の記録を含む電子データベースを提供し、さらに疾病または障害に関与する突然変異を有する患者において標的分子を決定する段階を含む。標的分子に結合しないか、またはその活性を調節しない治療薬が、データベースから選択される。1つの実施例において、突然変異によって、データベース内の1つ以上の治療薬に対する標的分子の親和性が減少し、その結果、変異のある被験者では、変異のない被験者に比べ、治療薬の有効性が減少することもある。本実施例によれば、標的分子以外の分子と結合する治療薬が選択される。他の実施例で、突然変異を有する被験者または被験者群は、標的分子の変異型に対し親和性が減少している治療薬の臨床試験から除外されるか、あるいはその被験者または被験者群は臨床試験の特定サブグループとして分類される。他の実施例で、突然変異を有する被験者または被験者群は、標的分子以外の分子と結合する治療薬の臨床試験で選択されるか、あるいはその被験者または被験者群は臨床試験の特定サブグループとして分類される。特定な実施例で、標的分子は、タンパク質または核酸である。1つの実施例で、この方法はコンピュータで実行されている。

30

40

【0056】

本発明はまた、関心対象の化合物が試料中に存在しているかを決定するための質量分析の改良方法の特徴とする。これらの方法を使用して、特定な標的分子に対するリガンドを同定し得る。

【0057】

上述のような1つの局面において、本発明は、関心対象の化合物が試料中に存在しているかを決定する方法を提供する。この方法は、(i) 化合物ライブラリから得た2つ以上の化合物の参照質量スペクトル、および(ii) そのライブラリから得た1つ以上の化合物を含む試料の被験質量スペクトルの決定または提供を含む。参照質量スペクトルの1つ以上のピークが被験質量スペクトルに含まれているかを決定し、その結果、参照質量スペクトルを

50

生成した化合物が試料中に存在しているかを決定する。様々な実施例において、被験質量スペクトルの全ピークが1つの化合物に割り当てられるまで、参照質量スペクトルを順にまたは同時に解析する。他の実施例では、参照質量スペクトルのピークが被験質量スペクトルに含まれているか否かを決定する段階は、1つ以上の参照質量スペクトルのピークが被験質量スペクトルに含まれているか否かを順番に決定する段階を含む。さらに他の実施例で、参照質量スペクトルのピークが被験質量スペクトルに含まれているか否かの決定は、(i) 参照質量スペクトルの全ピークが被験質量スペクトル中に存在すると決定し、それにより、参照質量スペクトルを生成した化合物が試料中に存在することを決定するまで、または(ii) 参照質量スペクトルの1つのピークが被験質量スペクトル中に存在しないと決定し、それにより、参照質量スペクトルを生成した化合物が試料中に存在しないと決定するまで、繰り返される。

10

【0058】

さらに別な局面において、本発明は、関心対象の化合物が試料中に存在しているかを決定する別法を提供する。この方法は、(i) 化合物ライブラリから得た2つ以上の化合物の参照質量スペクトル、および(ii) そのライブラリから得た1つ以上の化合物を含む試料の被験質量スペクトルの決定または提供を含む。被験質量スペクトルの1つ以上のピークを解析し、それらのピークが参照質量スペクトルに含まれているかを決定する。被験質量スペクトル中に存在する1つのピークを含む参照質量スペクトルに対して、参照質量スペクトルの1つ以上の他のピークを解析し、それらのピークが被験質量スペクトル中に存在するかを決定し、それにより、参照質量スペクトルを生成する化合物が試料中に存在するかを決定する。特定な実施例は、参照質量スペクトルのピークが被験質量スペクトル中に存在しているかを決定する段階は、1つ以上の参照質量スペクトルのピークが被験質量スペクトルに含まれているかを順番にまたは同時に決定する段階を含む。他の実施例で、参照質量スペクトルの1つのピークが被験質量スペクトル中に存在するかの決定は、(i) 参照質量スペクトルの全ピークが被験質量スペクトル中に存在することを決定し、それにより、参照質量スペクトルを生成した化合物が試料中に存在することを決定するまで、または(ii) 参照質量スペクトルの1つのピークが被験質量スペクトル中に存在しないと決定し、それにより、参照質量スペクトルを生成した化合物が試料中に存在しないと決定するまで、繰り返される。

20

【0059】

関心対象の化合物が試料中に存在するかを決定する上述のいずれかの方法についての様々な実施例において、ライブラリ中の各化合物質量スペクトルが決定される。さらに他の実施例では、参照スペクトルの少なくとも1つのピークは同位体ピーク、フラグメントピーク、または親ピークである。特定な実施例で、本方法は、参照スペクトルの全ピークが被験質量スペクトル中に存在しているかの決定を含む。他の実施例では、参照質量スペクトルは、質量スペクトルを生成する化合物の記録に関連する、質量スペクトルの1つ以上の特性の記録を含むデータベースに含まれている。特定な実施例において、データベースには、同位体ピークの質量/電荷比、フラグメントピークの質量/電荷比、親ピークの質量/電荷比、同位体のピーク強度、フラグメントのピーク強度、および親ピークの強度からなる群より選択される1つ以上の特性に関するデータが含まれる。さらに他の実施例で、1つの被験質量スペクトルピークが参照質量スペクトル中に存在しているかを決定する1つ以上の操作は、コンピュータによって実行されている。

30

40

【0060】

本発明においてはまた、関心対象の化合物が試料中に存在しているかを決定するプログラムを内蔵したコンピュータ読み取り可能なメモリを提供する。このコンピュータ読み取り可能なメモリは、参照質量スペクトル（すなわち、化合物のライブラリから得た個々の化合物の質量スペクトル）の1つ以上のピークの質量/電荷比を含む質量分析データを入力値として受け取るコンピュータコードを含む。このコンピュータ読み取り可能なメモリはまた、被験質量スペクトル（すなわち、ライブラリから得た1つ以上の化合物を含む試料の質量スペクトル）の1つ以上のピークの質量/電荷比を含む質量分析データを入力値として受

50

け取るコンピュータコードを含む。コンピュータ読み取り可能なメモリはまた、参照質量スペクトルのピークが被験質量スペクトルに含まれているかを決定し、それにより、参照質量スペクトルを生成する化合物が試料中に存在するかを決定する、コンピュータコードを有する。

【0061】

関連する局面において、発明は、関心対象の化合物が試料に存在しているかを決定するプログラムを内蔵したコンピュータ読み取り可能なメモリを特徴とする。そのメモリは、参照質量スペクトル（すなわち、化合物ライブラリから得た個々の化合物の質量スペクトル）の1つ以上のピークの質量/電荷比を含む質量分析データを入力値として受け取るコンピュータコード、および被験質量スペクトル（すなわち、ライブラリの1つ以上の化合物を含む試料の質量スペクトル）の1つ以上のピークの質量/電荷比を含む質量分析データを入力値として受け取るコンピュータコードを含む。またメモリは、被験質量スペクトルの1つ以上のピークが参照質量スペクトルに含まれているかを決定するコンピュータコード、および、参照質量スペクトルの全ピークが被験質量スペクトル中に存在しているかを決定し、それにより、参照質量スペクトルを生成した化合物が試料中に存在しているかを決定する、コンピュータコードも含む。

【0062】

本発明はまた、発現ベクター生産の自動化またはタンパク質の生産および生成の自動化方法の特徴とする。

【0063】

そのような1つの局面において、本発明は、関心対象のタンパク質をコードする2つ以上のベクターを作成する方法を特徴としている。この方法は、関心対象の第一タンパク質をコードする第一核酸と第一骨格核酸を、それらが反応し得る条件下で、ロボット操作装置を使用しロボット操作で接触させ、それにより、第一タンパク質をコードする第一ベクターを作成する段階、さらに、関心対象の第二タンパク質をコードする第二核酸と第二ベクター核酸を、それらが反応し得る条件下で、ロボット操作装置を使用しロボット操作で接触させ、それにより、第二タンパク質をコードする第二ベクターを作成する段階に關与する。いくつかの実施例では、本法はまた、第一ベクターを第一細胞に挿入し得る条件下で、第一ベクターを第一細胞にロボット操作で接触させる段階、ならびに第二ベクターを第二細胞に挿入し得る条件下で、第二ベクターを第二細胞にロボット操作で接触させる段階を含む。様々な実施例で、少なくとも3、4、5、8、10、15、30、60、90個またはそれ以上のベクターが同時に作製される。他の実施例では、骨格核酸は発現ベクターに線形化され、関心対象のタンパク質をコードする挿入片は、挿入片を含む円形ベクターを生成し得る条件下で、発現ベクターと連結される。他の実施例では、第一と第二ベクターまたは細胞はロボット操作装置の別々のフラスコまたはウェルに入れられる。他の実施例では、第一細胞は第一タンパク質を発現し、第二細胞は第二タンパク質を発現する。さらに他の実施例では、第一タンパク質と第二タンパク質は、以下の局面記載のように精製される。他の実施例で、第一細胞/第二細胞は、大腸菌のような細菌、ショジョウバエ細胞のような昆虫細胞、またはCos、HEK293、またはCHOのような哺乳類の細胞である。他の実施例で、第一ベクターと第二ベクターは、第一タンパク質と第二タンパク質の生成のため、第一細胞と第二細胞から昆虫や哺乳類細胞のような別の細胞種に移される。他の実施例において、回転ボトルシステム、Stirタンクシステム、キャピラリー細胞培養システム、またはバイオリアクターを使って細胞が増殖される。第一ベクターおよび/または第二ベクターを使って、本発明のいずれの方法（例えば、タンパク質に結合するリガンドを同定する）でも使用されるタンパク質を生成できる。

【0064】

本発明の1つのタンパク質の生成および/または精製方法は、結果としてロボット操作装置内の第一培養液中に第一タンパク質が分泌される条件下で、第一タンパク質を第一細胞中で発現すること、また結果としてロボット操作装置内の第二培養液中に第二タンパク質を分泌する条件下で、第二タンパク質を第二細胞中で発現することを含む。ロボット操作装

置により、第一培養液が第一クロマトグラフィーのカラムに移され、第二培養液が第二クロマトグラフィーのカラムに移される。1つの実施例で、第一タンパク質と第二タンパク質が単離され、それにより第一タンパク質および第二タンパク質を精製する。様々な実施例で、少なくとも3、4、5、8、10、15、30、60、90個またはそれ以上のタンパク質が同時に精製される。他の実施例では、第一と第二細胞はロボット操作装置の別々のフラスコまたはウェルに入れられる。他の実施例では、第一細胞/第二細胞は、大腸菌のような細菌、ショジョウバエ細胞のような昆虫細胞、またはCos、HEK293、またはCHOのような哺乳類の細胞である。他の実施例で、第一細胞および/または第二細胞は、一過的に形質移入されたCos、HEK293、ショウジョウバエ細胞、またはCHO細胞、あるいは安定して形質移入されたCos、HEK293、CHO、大腸菌、またはショウジョウバエ細胞である。さらに他の実施例で、第一タンパク質および/または第二タンパク質は哺乳類または昆虫の細胞中でグルコシル化される。様々な実施例において、第一タンパク質または第二タンパク質は、天然に分泌シグナルを有しているか、または分泌シグナルを持つように遺伝的に修飾されており、その結果、それらのタンパク質は細胞から培養液中に分泌される。第一タンパク質および/または第二タンパク質は、本発明のいずれの方法（例えば、タンパク質に結合するリガンドを同定する）においても使用できる。他の実施例で、ロボット操作装置を使用して、第一タンパク質および/または第二タンパク質を候補リガンドのライブラリと接触させ、本書で記載されているいずれかの方法によって、タンパク質と結合するリガンドを選択できる。さらに他の実施例では、関心対象の小分子とロボット操作で接触する標的分子ライブラリのメンバーとして、第一タンパク質および/または第二タンパク質を使用し、

10

20

【0065】

本発明のいずれかの局面の様々な態様において、リガンドは共有結合または非共有結合で標的分子と結合する。他の実施例で、リガンドは、標的分子と結合するかまたは標的分子と同じ経路内の別の分子と結合し、それにより、標的分子を活性化または阻害する。他の実施例で、リガンドの分子量は、5000、4000、3000、2000、1000、750、500または250 ダルトン未満である。他の実施例で、リガンドは、5、4、3、または2個未満の水素結合ドナーを、または10、8、6、4、または3個の水素結合ドナーを有する。さらに他の実施例で、リガンドの $c \log P$ は4.15未満である。さらに他の実施例で、リガンドはFK506ではない。他の実施例で、選択された候補リガンドは、 K_d が1 fM未満、1 fMから1 nMの間、1 nMから1 μ Mの間、または1 μ M未満で標的分子と結合する。他の実施例で、選択された候補リガンドは、IR、MS、NMR、UV、アミノ酸配列決定、核酸配列決定、またはそれらを併用した解析を受ける。他の実施例で、同位体またはフラグメントピークを使用し、ライブラリの別の候補リガンドと同じ質量を持つ候補リガンドを同定する。

30

【0066】

本発明のいずれの局面の様々な他の態様において、候補リガンドおよび/または標的分子は溶液相に存在している。他の実施例では、リガンドまたは標的分子はビーズまたはチップのような固体表面に固定化される。他の実施例では、アッセイ培養液はクロマトグラフィーによって画分に分別される。特定の実施例では、複合体は、サイズ排除型（例えば、シリカ、またはポリマー樹脂を使用）、多モード型、二モード型、または二相型クロマトグラフィー（例えば、サイズ排除型の逆相、サイズ排除型のアニオン交換、サイズ排除型のカチオン交換など2つ以上の特徴を有するクロマトグラフィー、または内部表面逆相（ISRP）、GFFあるいはGFFII樹脂）によって、単離される。代表的な樹脂は、ジオール、セファロース、スペロース、およびポリメチルメタクリレートを含む。他の好ましい樹脂は、5、50、500、500、5000、あるいは7000 psiを超えても安定である。特定の実施例で、異なる分離特性を持つ樹脂を含むカラムが順番に組み合わされる。他の実施例では、カラムクロマトグラフィーが複合体の単離に使用され、複合体は60、30、20、15、10、5、3、2、または1分未満でカラムから溶出し、ボイド量は20、15、10、5、4、3、2、または1 mL未満であり、あるいは、カラム直径は、5、4、3、2、または1 mm未満である。他の実施例

40

50

では、HPLC、スピンカラム、キャピラリークロマトグラフィーまたはろ過を使用して、複合体を単離する。他の実施例では、HPLCのUV吸収の減少または非結合リガンドに相当する他のクロマトグラフィーピークを使用し、非結合リガンド量の減少を検出している（つまり、結合リガンド量の増加）。さらに他の実施例では、標的分子と結合候補リガンドの複合体には、結合リガンドを標的分子から分離するためのクロマトグラフィー操作を行う。本発明のいずれかの局面のまた他の実施例において、固定化された標的は、候補リガンドと接触され、さらに支持体は候補リガンド不含の培養液で洗浄され、すべての結合リガンドを標的から遊離する方法で処理される。さらに他の実施例では、標的を候補リガンドに曝露後、支持体を標的分子不含の培養液で洗浄し、支持体から候補リガンド分子およびあらゆる結合標的分子を取り除く方法で、支持対を処理する。他の局面において、この方法の1つ、複数またはすべての操作がロボット操作で自動化されるか、またはコンピュータで実行されている。

【0067】

本発明のあらゆる局面のさらに他の態様において、選択された標的の機能または活性は、化学的アッセイ、生化学的アッセイ、酵素アッセイ、生物学的アッセイ、またはそれらの組み合わせによって特徴づけられる。特定な実施例で、標的分子の機能は、アポトーシスアッセイ、増殖アッセイ、壊死アッセイ、血管形成アッセイ、浸潤アッセイ、またはそれらを併用して特徴づけられる。他の実施例で、候補標的分子は、生化学的抽出物、細胞、組織、生物体、または遺伝子組み換え源から単離される。さらに他の実施例で、選択された標的分子は、NMR、IR、UV、MS（例えば MALDITOF、MALDI、シングル4重極、トリプル4重極、または電子スプレイ、MSまたはMS-MS（タンデムマス））、アミノ酸配列決定法、または核酸配列決定法によって、同定される。他の実施例では、候補標的分子は全長タンパク質または全長に足りないタンパク質のフラグメントである。代表的な標的には、GPCR、キナーゼ、イオンチャネル、核レセプター、プロテアーゼ、ホスファターゼ、およびメチラーゼのような酵素およびレセプターが含まれる。標的には、治療薬として有効な化合物が以前に開発された、またはされていない分子または分子のクラスを含むこともある。

【0068】

候補リガンドについての本発明の様々な局面の全態様を、関心対象の小分子に適用することが指摘されている。

【0069】

ここで、「薬物標的として以前に確認されていない標的分子」とは、疾病の動物モデルで疾病状態を促進または阻害するための調節が、出版または公開發表の記述として、以前に実験的に測定されていない標的分子を意味する。例えば、未確認の標的分子には、分子の活性化または阻害、あるいは分子発現レベルの減少または増加が、動物モデルの疾病状態を調節することが、実験的に示されていない分子が含まれる。これに対し、確認済みの薬物標的には、分子の量の増減または分子活性が、動物モデルの疾病状態を促進するか阻害するかが実験的に証明されている分子が含まれる。確認済みの標的例には、ノックアウト突然変異またはその他の遺伝子サイレンス方法（例えば、遺伝子発現のアンチセンス阻害）による過剰発現あるいは不活性化が、動物モデルの疾病状態を促進または阻害するか、実験的に明らかにされている標的が含まれる。

【0070】

「生物学的機能が未知の標的分子」とは、その活性が、出版または公開發表の記述として、以前に実験的に証明されていない標的分子を意味する。様々な実施例で、機能未知の標的分子とは、活性が実験的にすでに証明されている核酸またはタンパク質への配列同一性が60、50、40、30、20、または10%未満である核酸またはタンパク質のことである。他の実施例では、それらの核酸またはタンパク質は推定機能が以前に与えられていない。配列の同一性は、通常は、その中で指定された初期設定のパラメータで配列解析ソフトウェアを行い測定される（例えば、Genetics Computer GroupのSequence Analysis Software Package、University of Wisconsin Biotechnology Center, 1710 University Avenue, Madison, WI 53705）。このソフトウェアプログラムは、色々な置換、削除および他の変更に

10

20

30

40

50

対する相同性の程度を与えることによって、類似配列と一致させる。

【0071】

「二次または三次構造が未知の標的分子」とは、二次または三次構造が、出版または公開発表の記述として、以前に実験的に決定されていない標的分子を意味する。いくつかの実施例で、二次または三次構造は、相同性分子の既知構造に基づいて以前に予測またはモデル化されていない。他の実施例で、標的分子の結合部位あるいは活性部位の場所または三次構造は、以前に実験的に決定されていない。

【0072】

「骨格」とは、候補化合物ライブラリ中の2つ以上の異なる分子が含まれる中心化学構造を意味する。様々な実施例では、ライブラリの少なくとも5、10、 10^2 、 10^3 、 10^4 、 10^5 、 10^6 個またはそれ以上の分子が骨格を含んでいる。いくつかの実施例では、ライブラリは、少なくとも2、2、5、10、 10^2 、 10^3 、 10^4 、 10^5 種またはそれ以上の異なる骨格を含んでいる。

【0073】

「ライブラリ」とは、またはそれ以上の異なる分子の集まりを意味する。2、5、10、 10^2 、 10^3 、 10^4 、 10^5 、 10^6 、 10^7 、 10^8 、 10^9 個またはそれ以上の異なる分子の収集を意味する。様々な実施例において、各ライブラリのメンバーは異なる質量を有する。他の実施例では、少なくとも2、5、10、15、20、30、40、50個またはそれ以上のメンバーが、他のライブラリメンバーと同じ質量か、あるいはその差が1、0.5、0.1、0.05または0.01ダルトン未満の質量を有する。

【0074】

「プロテオーム」とは、生物体によって発現されるすべてのタンパク質を意味する。プロテオームは、生物体によって発現されるタンパク質の、すべての選択的スプライシング変異体を含む。

【0075】

「精製された」とは、天然に共存する他の成分から分離することを意味する。一般的に、化合物は、少なくとも重量の50%に、タンパク質、抗体および天然に会合する天然有機分子を含んでいない場合、実質的に純粋である。他の実施例では、化合物は、少なくとも重量で75%、90%または99%が純粋である。実質的に純粋な化合物は、化学的合成、天然源からの化合物の分離、またはその化合物を天然に生成しない遺伝子組み換え宿主細胞中での化合物の生成によって得られ得る。タンパク質および有機化合物は、Ausubelらによって記述されたような標準技法を用い、熟練者により精製され得る。(Current Protocols in Molecular Biology, John Wiley & Sons, New York, 2000)。出発物質に比した精製度は、ポリアクリルアミドゲル電気泳動、カラムクロマトグラフィー、光学密度、HPLC分析、またはウェスタン分析(Ausubelら)のような標準方法によって測定できる。代表的な精製法には、免疫沈降、免疫親和性のようなカラムクロマトグラフィー、磁気ビーズ免疫親和性精製、およびプレート結合抗体によるパニングなどが含まれる。

【0076】

本発明の方法は多数の利点を有している。例えば、本方法によって、生物体のプロテオーム(例えば、ヒトプロテオーム)のすべてのタンパク質の発現と精製、各タンパク質の高親和性薬物様骨格の同定ができる。本方法はまた、理論的には無制限に、候補化合物および候補骨格物をスクリーニングできる。本発明の方法は、迅速にまた大規模スケールで実施できるので、以前に薬物標的として確認されていない標的分子または生物学的機能が未知の標的分子をアッセイし、標的分子に結合および/またはその活性を調節するリガンドを選択するのに有用である。これに対し、標的分子と結合するリガンドを選択するための現行方法は、薬物標的として確認されている標的分子だけに限定されている。そのため、本方法はアッセイされる標的分子数を非常に拡大している。高親和性結合剤が選択される標的分子は、それで、薬物標的として確認することができる。

【0077】

さらに、本発明の方法によって、同じ質量を持つ候補リガンドを区別することができる。

例えば、質量スペクトル同位体およびフラグメントのピークは、一般的に同じ質量のリガンドであっても異なる。そのため、候補リガンドが化合物ライブラリの別の候補リガンドと同一の親ピークを有していたとしても、これらのピークを利用して候補リガンドを同定できる。この利点により、同一または類似質量数の複数化合物を含むライブラリを使用することができる。

【0078】

本発明の溶液相の態様により、液相の結合を、血清または細胞中で起きていると思われる状態で生じさせることができる。標的タンパク質の特異的な作用を測定する多数の現行方法に比べ、本発明の方法は、カスタマイズせずにプロテオーム中のどの標的にもすぐに適用され得る。本方法はまた、非常に少量の試薬しか消費しない（200,000 化合物に対する各標的量は<300 ug であり、各標的に対する各化合物量は<35 ng である）。本方法により、スクリーニング前に、個々のライブラリメンバーをタギングまたは精製せずに、化合物のライブラリをスクリーニングでき、それにより、ライブラリのスクリーニングに要する時間を大幅に短縮できる。ライブラリのスクリーニングに要する時間もまた、複数のライブラリおよび/または複数の標的を同時に解析できる、本発明の自動化実施例によって短縮できる。

【0079】

本発明の他の利点及び態様は、以下の詳細な説明と特許請求の範囲から明らかである。

【0080】

5. 発明の詳細な説明

5.1. 遺伝子型から表現型へ

1つの局面において、本発明は、タンパク質または核酸標的を多数の可能なリガンドに曝露させ、リガンド-標的対を収集し、そして標的と結合するリガンドを使用して標的物質の生物学的機能を分析する方法に関連する。1つの実施例を図1に概略する。この方法は、今まで知られていない標的物質の機能を決定するために使用され得る。標的分子に結合する候補リガンドを選択する他の多数の方法が本書に記載されている。セクション5.1.1〜5.1.5の下にリストされた全実施例は、本発明のいずれの方法でも使用できる。

【0081】

5.1.1. 標的物質

本発明によれば、標的分子とは、それに結合または反応する分子が模索されている化合物である。好ましい実施例では、標的は、反応容器中の最大の濃度で存在する物質である。様々な好ましい実施例では、標的は、反応容器のリガンドと同じ濃度で存在している。さらに他の好ましい実施例で、その標的は、各リガンド濃度または、候補リガンドの全混合物濃度よりも高いかまたは低い。他の好ましい実施例で、標的は、反応容器中の最小濃度で存在する物質である。発明の1つの実施例で、標的は、反応容器中の最大の分子量を有する物質である。標的は、インビボまたはインビトロで合成された天然に存在する生体分子であってもよい。標的は、アミノ酸、核酸、糖、脂質、天然物質またはそれらの組み合わせでもよい。instant発明の利点は、標的物質の正体または機能の予備知識を必要としないことである。

【0082】

本発明の好ましい態様では、標的とは、アミノ酸、ペプチド、酵素、タンパク質、抗体、またはそれらの組み合わせから構成される。第一段階は、対象タンパク質をコードするポリヌクレオチドを選択し、発現系に導入することである。ポリヌクレオチドは、示差スクリーニング、差し引きハイブリダイゼーション、ディファレンシャル ディスプレイ（示差表示）、マイクロアレイ発現解析、発現差解析（RDA）、またはレーザー捕獲顕微解剖によって選択され得る。タンパク質は、細菌プラスミド、ファージ、一過性の細胞発現系、またはウイルス発現系としてインビボで合成される。また、選択されたタンパク質はインビトロの転写および翻訳（例えば、Promegaウェブサイト）、あるいは一般的なFMOCオリゴペプチド合成化学によってインビトロで合成されることもある。発現されたタンパク質は、随意に精製され、続いてリガンドライブラリに曝露される。

【0083】

本発明によれば、遺伝子は、ヒトあるいは他種の完全なcDNAまたは遺伝子ライブラリ、あるいは特定の疾患あるいは特定の刺激物中で識別発現用に選択した遺伝子のサブセットから発現され得る。疾病または刺激された細胞および組織中で識別的に発現される遺伝子は、差し引きハイブリダイゼーション、情報学、マイクロアレイ、SAGE、またはレーザー捕獲顕微解剖のような技法、またこれだけに限定されないが、これらの方法によって選択され得る。ESTのような部分的配列が回収されると、全長組織特異性cDNAsが全長ヒトcDNAライブラリからクローン化されることもあり、それらのいくつかは、CLONTECH、STRATAGENE、Life Technologies社、およびNCBIなどから入手可能である。この方法でクローン化された遺伝子の20%から60%は、組織にも依存するが、これまで同定されておらず、クローン化された実際にすべての遺伝子の機能は解明されていない。好ましい実施例において、これらの遺伝子はゲノミクスによって発見された。タンパク質を生成するために、cDNAの全長が、遺伝子のカルボキシ末端に挿入されたヘキサヒスチジン(6his)、およびアミノ末端のグルタチオンシンセターゼ(GST)によってタグが付けられる(各々にプロテアーゼサイトが存在)。また、プロテアーゼ処理を避けるために、New England Biolabs社によるインテインに基づく自己開裂タグも使用できる。これらの遺伝子は、発現され、バキュロウィスルの上澄み液に分泌される、例えば、HisタグとBIPタンパク質リーダーを伴うInvitrogen社ショウジョウバエSchneider 2細胞系、CaHPO₄による形質移入、および硫酸銅を用いるヒグロミシン誘発発現による選択などによって起こり、これにより、ニッケルカラムで精製される上澄み中には、5~10 mg/Lのタンパク質が生成され得る。代替発現系の限定されない例には、Fast Bacまたは別のバキュロウィルス系、あるいは哺乳類発現系(CH0、COS、293など)が含まれる。大腸菌はタンパク質生成に使用されるが、タンパク質をグリコシル化せず、またバキュロウィルスは同様に信頼性があり、確実にタンパク質をグリコシル化する。得られたタンパク質は続いて、精製の第一段階としてNi(2+)-NTAクロマトグラフィーによって、第二段階としてグルタチオン親和性クロマトグラフィーによって精製が可能で、次にタグの開裂による特異的なプロテアーゼ除去が行われる。インテインに基づく親和性システムが使用される場合、プロテアーゼは必要はない。タンパク質は代替技法によって同様に発現および精製でき、あるいは全長または部分的なタンパク質がファージ中に発現または表面に結合される。

【0084】

本発明の別の態様では、標的物質はオリゴヌクレオチドまたはポリヌクレオチドとして、RNAまたはDNAから構成される。本発明の限定されない1つの実施例において、発現系内に導入される核酸は、ESTの大規模配列決定によって同定される。オリゴヌクレオチド標的は、直接合成され得る。ポリヌクレオチド標的は直接合成されるか、またはテンプレートポリヌクレオチドの増幅、例えばPCRにより、作製され得る。オリゴヌクレオチドあるいはポリヌクレオチド標的は、随意に精製され、続いてリガンドライブラリに曝露される。

【0085】

本発明の別の態様では、標的は単純または複合糖質から構成される。本発明の別の実施例では、標的は脂質から構成される。本発明の別の実施例では、標的は天然物から構成される。

【0086】

本発明の別の態様では、標的化合物は誘導体化されることもある。限定されない例として、ビオチン、フルオレセイン、ジゴキシゲニン、緑蛍光タンパク質、放射性同位体、ヒスチジンタグ、磁気ビーズ、グルタチオンSトランスフェラーゼ、光活性化架橋剤あるいはそれらの組み合わせが含まれる。

【0087】

標的物質の調製において、必要なコンポーネントの部分的または不完全な精製の結果、少量の他の化合物が含まれることもある。

【0088】

5.1.2. リガンド

本発明によれば、リガンドは、ターゲットに結合するか、生物学的アッセイで効果を発揮するか、の両方または一方の可能性のある分子である。遺伝子型から表現型へのアプローチの様々な実施例において、リガンドまたは候補リガンドの混合物の濃度は、反応容器中でターゲット濃度よりも低い。遺伝子型から表現型へのアプローチの他の実施例において、リガンドまたは候補リガンドの混合物の濃度は、反応容器中でターゲット濃度と同じである。遺伝子型から表現型へのアプローチのさらに他の実施例において、リガンドまたは候補リガンドの混合物の濃度は、反応容器中でターゲット濃度よりも高い。標的は、アミノ酸、核酸、糖質、脂質、天然物質、天然物質様化合物またはそれらの組み合わせでもよい。リガンドは、化学的方法のいずれの組み合わせからも作製され得る。また、リガンドは天然に存在する生体分子をインビボまたはインビトロで合成された物かもしれない。リガンドは、別の化合物から随意に誘導体化され得る。この修正の1つの利点は、誘導体化する化合物を、リガンド-標的複合体の収集またはリガンドの収集を容易にするために、例えば、リガンドと標的を分離した後に、使用できることである。誘導体化グループの限定されない試料には、ビオチン、フルオレセイン、ジゴキシゲニン、緑色蛍光タンパク質、放射性同位体、ポリヒスチジン、磁気ビーズ、グルタチオンS トランスフェラーゼ、光活性化架橋剤あるいはそれらの組み合わせが含まれる。

【0089】

リガンドは、標的がリガンドライブラリに曝露される条件下で、お互いの親和性は低くなければならない。

【0090】

リガンドライブラリは、質量、組成、構造またはその組み合わせが異なるリガンド混合物である。本発明では、少なくとも10個の異なるリガンド、あるいは少なくとも100個の異なるリガンド、また少なくとも1000個の異なるリガンドを含むライブラリを意図している。

【0091】

タンパク質に結合するリガンドライブラリは、多数のソース物質から誘導され得る。本発明では、化学物質、タンパク質、ペプチド、抗体、糖質、脂質、天然物質、天然物質様化合物またはそれらの組み合わせを使用している。これらの物質は、有機合成、組み合わせ化学、組み換えDNA、生化学抽出、精製などによって、調製される。本発明の好ましい実施例では、天然物質様合成ライブラリは、多様な化学的方法（例えば、ビーズ上または溶液における非対称分割プール合成、同時あるいは連続合成など）、また組み合わせ化学あるいは医薬品学などを駆使して生成される。合成で使用されるサブユニットは、できれば薬物類似であり、できるだけ高く多様化している。このユニットは、構造的に固定されていても柔軟性があってもよい。ユニットはその構造を修正する化学的応答を受ける（例えば、転位など）。ユニットに官能基が追加されることもある。

【0092】

薬物様化合物は、異なる化学方法を用い、異なる骨格を使用して作製されるであろう（例えば、有機、無機、ペプチド、タンパク質、アルカロイド、糖質、脂質、天然物様化合物など）。薬物様化合物はスペクトル上の識別子を有していることがある。スペクトル識別子の限定のない例として、質量分析上で特徴的な同位体分裂パターンに分解する元素が挙げられる（例えば、Cl、Br、N、H）。薬物様化合物は、質量分析で特徴的な分裂パターンを持つ化合物から作製され得る（ペニシリン）。ライブラリを、他の分析法および逆重量積分法（例えば、IR FTIR）で容易に測定できるようにデザインすることができる。

【0093】

本発明の別の態様で、使用される可能性のある他のライブラリの限定されない例には、市販のライブラリも含まれ（例えば、Pharmacopeia、ArQule、およびChembridge）、特殊な化学ライブラリ、ペプチド、TAT、VP22またはANTENNAPEDIA形質移入シグナルなどのペプチドまたはタンパク質、構造的に柔軟性のある小分子、天然物質、糖質、およびモノクローナル抗体などが含まれる。合成で使用されるサブユニットは、できれば薬物類似であり、できるだけ高く多様化している。

【0094】

本発明のライブラリには、結合が観察された後、リガンド逆重量積分法および再合成を容易にするためタグが付けられ得る。また、リガンドはタギングなしで逆重量積分できる。リガンドは個別にまたは混合物として測定できる。液相または固相支持体での混合物として合成された多様なライブラリが使用できる。1つの実施例で、形質導入ペプチドまたはTAT、VP22またはANTENNAPEDIAの変異体は、小分子と架橋し、細胞膜あるいはバリアーを横断する力が増強される。また、これらのペプチドの小分子同族体が開発され、同じ物質にリンクされ得る。

【0095】

5.1.3. 結合

本発明によれば、1対のリガンド-標的対を用いて、解離定数 (K_d) が $20 \mu\text{M}$ 未満、好ましくは約 $1 \mu\text{M}$ 未満のリガンドと標的間の親和性を説明している。本発明はさらに、 $K_d \leq 100 \text{ nM}$ 、 $K_d \leq 100 \text{ pM}$ 、 $K_d \leq 100 \text{ fM}$ のリガンド-標的相互作用を予期している。これらの相互作用は共有結合性または、非共有結合性である。リガンド-標的対のリガンドは、他の標的に親和性を示すことも示さないこともある。リガンド-標的対の標的は、他のリガンドに親和性を示すことも示さないこともある。

【0096】

本発明によれば、反応容器とは、標的が少なくとも1つのリガンドに曝露されると思われる容器中あるいは表面上のことである。本発明の好ましい実施例では、反応容器は、ハイスループットスクリーニングを容易に行えるように用意されている。このスクリーニングは、96個または384個のウェルを持つマイクロタイター用プレート上で実施される。別の可能性は、MacBeathらが示した、ガラススライド上で異なる標的タンパク質を高濃度に配置させる方法である (MacBeath et al., 2000, Science 289:1760)。本発明の他の実施例で、反応容器はカラム、樹脂、膜、マトリックス、ビーズまたはチップである。

【0097】

標的がリガンドライブラリに曝露される条件は様々である。限定のない例では、結合反応の温度は、約 5°C 未満、約 5°C から約 25°C 、約 25°C から約 40°C 、あるいは 40°C を超える。さらに、限定のない例で、結合反応の条件は、約 pH 5 未満、約 pH 5 から約 9、あるいは約 pH 9 を超える。さらに限定のない例で、結合反応の溶液は、水、アルコール、有機溶媒、またはそれらの組み合わせである。さらに限定のない例で、結合反応の条件として、添加剤はイオン、塩、洗剤、還元剤、酸化剤、またはそれらの組み合わせである。さらに限定のない例では、結合反応の条件として標的は固定化される。さらに限定のない例では、結合反応の条件としてリガンドは固定化される。さらに限定のない例では、結合反応の条件として複数の標的は固定化される。さらに限定のない例では、結合反応の条件として標的とリガンドは溶液状態である。

【0098】

さらに非限定的な実施例において、結合反応の条件として、リガンドはビオチン、フルオレセイン、ジゴキシゲニン、緑蛍光タンパク質、放射性同位体、His (ヒスチジン) タグ、磁気ビーズ、酵素またはこれらの組み合わせを構成する。

【0099】

本発明の1つの態様で、標的は機械装置に基づくアッセイ法でスクリーニングされる。機械装置に基づくアッセイは、標的と結合するリガンドを検出するアッセイであるが、それだけに限られない。このアッセイは、リガンド、タンパク質、または検出の指標となるいずれかの物質と固相または液相で結合するイベントを含む。また、機能が今までに確認されていないタンパク質をコードする遺伝子は、レポーター系 (β -ガラクトシダーゼ、ルシフェラーゼ、緑色蛍光タンパク質など、しかしこれだけに限られない) によって細胞内に形質移入され、理想的にはハイスループットや超ハイスループット (例えば、チッププレート1枚につき1560個のウェル) スクリーニング法、またはライブラリの個々のメンバーによって、ライブラリに対してスクリーニングされる。本発明の別の実施例では、他の機械装置に基づく結合アッセイが使用され得る。これらのアッセイには、酵素活性への

10

20

30

40

50

作用を測定する生化学的アッセイ、標的およびレポーターシステム（例えば、ルシフェラーゼまたは β -ガラクトシダーゼ）が1つの細胞に導入された細胞に基づくアッセイ、または自由エネルギーの変化を検出する結合アッセイなどの他のアッセイが入る。結合アッセイはウェル、ビーズ、またはチップに固定されるか、固定化抗体に捕獲されるか、またはキャピラリー電気泳動によって分離された標的を用いて実行できる。結合したリガンドは、比色分析、蛍光分析、あるいは表面プラズモン共鳴法によって通常検出される。カラムに基づく結合アッセイにおいて、結合はウェルまたは他の容器、ゲル上などで実施される。

【0100】

これらのアッセイを行う多数の手段があるが、帰納的思考してみると、標的タンパク質に結合する化学物質のみが関連を持ちその機能を伝えることができる。さらに、液相が真の生体構造をより正確に反映している。さらに、反応において、タンパク質および化学物質ともタグされないのが好ましく、その結果、タンパク質がビーズプレートにカップリングすることにより、なんらかの様式で束縛されたり、またそのリガンドが、細胞や血液中に存在するであろう同じ液相で確認されないというような問題が減少する。結果として、本発明の好ましい実施例では、小容量（0.1 μ L～100 μ Lの好ましい範囲で1 fL～1 mL）中に、1～20,000個のリガンド（1000～10,000が好ましい）が1 ng～1 mgの各タンパク質（0.1 μ g～100 μ gが好ましい）に混合され、濃度は0.1 μ M～100 μ Mとなり、好ましい範囲は0.1 μ Mから10 μ Mである。本発明の特定な実施例において、各タンパク質にマイクロモルからナノモル単位の親和力で結合することが期待される、1～500個のリガンドのみに注目すれば、数百万種の組み合わせを個別にスクリーニングしなくてもすむことになる。これにより、分子自身の質量、同位体パターンまたは分解パターン以外の方法でライブラリをタグする必要がなくなる。なぜなら、質量分析法では、1つのウェルに付き1から5個のヒット化合物を分解し、同定する段階ができるからである。また、IRとFTIRの両方または一方を単独で、または質量分析と併用して、ヒット化合物を分解・同定できる。

【0101】

5.1.4. リガンド-標的対の分離とリガンドの同定

本発明の好ましい実施例において、リガンド-標的対は液体クロマトグラフィーによって、非結合リガンドおよび非結合標的から分離され、次いで、リガンド-標的対は2段階目の液体クロマトグラフィーによって各対に分離され、さらに結合しているリガンドが質量分析によって同定される。本発明の様々な実施例において、溶液相中の結合は、ウェル、試験管、またはカラム内で起こり得る。キャピラリー電気泳動、および/または他の検出方法が、ライブラリからリガンドを逆重畳積分するために使用され得る。特に、HPLC（高速液体クロマトグラフィー）と質量分析またはキャピラリー電気泳動と質量分析で、非常に高感度で分子を測定できる。さらに、この技法は、化学ライブラリの少量の各メンバーを至適に利用するのに重要である、極少量の使用で実行できる。例えば、化学ライブラリの20,000未満のリガンドは、96個のウェルプレートの各ウェル内で、約100 μ Lでタンパク質1 μ g中10 μ M以下の濃度で、再び結合するために、タンパク質とプールされ得る。好ましい実施例で、各ウェルのカラムとして作動するために、HPLCはカートリッジ付き96（個）ウェルプレートで実行される。別の実施例で、分離は、カラム、ウェル、カートリッジ、チップ、またはフィルタを用い、384個、1536個、あるいは10,000個以上のウェル型で同時に実行される。また、この分離は、標準のHPLCカラム、スピнкаラム、または他のカラムで実行され得る。最初のカートリッジ/カラムは、樹脂中の非結合分子を保持するが、結合リガンドとタンパク質を通過させるために、ゲル透過あるいはサイズ排除、またはゲル濾過型（例えば、G25様樹脂、Pharmacia）であり得る。少量の試料量が望まれ（好ましくは1～100 μ L以下）るが、さらにこの手順で試料は1桁以上希釈され得る。その故、試料の希釈を最小にするため、好ましくは、直径1～2 mm以下、長さ5～200 mm（Rocket Column、BioradまたはPharmacia社製カラム）の、小さくて径の狭いカラムを使用することが有用である。キャピラリー液体クロマトグラフィーも使用できる。この樹

脂によって、高い親和性で ($K_d \leq 1.0 \mu M$) で結合している小分子を伴ったタンパク質が分離される。次のカートリッジ／カラムでは、疎水性または親水性逆相HPLC樹脂を使用するであろう。使用するリガンドライブラリの疎水性によって以下のような樹脂を選択する。すなわち、C18（シリカ疎水性-疎水性が低いリガンド用）、C8カラム（より親水性で、より疎水性なりガンド用）、シアノカラム（より親水性なりガンド用）または、親水性あるいは疎水性リガンドどちらにも使用できるAgilent社SB8Uである。これらの逆相HPLCで、タンパク質に結合している小分子リガンドをタンパク質から分離し、小分子とタンパク質試料を樹脂結合を通じて濃縮する。続いて、小分子はタンパク質と樹脂から溶出され、溶出物は96（個）ウェルプレートで収集され得る。開始物質の量が既知であれば、この操作で親和性も測定し得る。また、後に、結合親和力を測定するために競合試験を実行できる。

10

【0102】

これらの溶出物は、質量分析測定に回され、特性が決定され得る。この測定は、96（個）ウェル型であっても、多分、同時多チャネルマイクロチップシステムまたは同時スプレイインタフェースのいずれかを用いて、リアルタイムでロボット操作で実施される可能性があり得る。また、チップに基づくMALDI（Matrix Assisted Laser Desorption Ionization）TOF（time of flight）質量分析も使用され得る。この場合、カラム（スピン、HPLC、キャプラリー、その他）から分離されたタンパク質画分は、96個ウェル以上のアレイ型のチップまたはフィルタの1つにスポット可能である。Bruker Daltonics社のOmniflexまたはAutoflex MALDI機によって、試料100個と1536個の試料型から、各試料が自動的に脱着し分析される。使用され得る限定のない型の質量分析として、電子スプレイ型、イオントラップ型、フーリエ変換型、MALDI型、シングルMS（Mass Spectroscopy）、MS-MS（タンデムマス）、あるいはMS-MS-MS（トリプルマス）型におけるシングルあるいはトリプル4重極であり得る。

20

【0103】

溶出物は、使用するリガンドライブラリの情報を補充した質量分析計と併用されるソフトウェアパッケージを使用して、特性が決定され得る。質量分析法は、直接質量を測定することにより、化合物を同定するために使用され得る。しかし、また、質量分析は、化合物、特徴的な同位体パターン（例えば、 ^{35}Cl , ^{13}N , ^2H ）に分解する元素を含む骨格またはリンカー、または特有な分裂パターンを有する化合物（例えばペニシリン）を検出するために使用され得る。例えば、塩素含有化合物は、3:1の強度比を有し2 AMU 離れている2つの質量ピークを生じる、 ^{35}Cl と ^{37}Cl から構成される。同様に、臭素含有化合物は、1:1の強度比を有し2 AMU 離れている2つの質量ピークを生じる、 ^{79}Br と ^{81}Br から構成される。このアプローチは、化合物を同定する際、代替え法として、または真の分子量と併用して使用され得る。

30

【0104】

質量分析は質量、同位体、および分解パターンを正確に測定できるので、ソフトウェアと併用すると、異性体を除くライブラリの正確なメンバーが同定され得る。この操作の後、理論的に予測される500個前後のマイクロモルあるいはナノモル濃度のヒット分子を、元のライブラリから引き出し、大規模スケールで合成することができる。その分子がペプチドである場合、タンパク質が細胞膜を横断できるようなTAT形質移入配列に融合させることができる。

40

【0105】

本発明の別の態様で、リガンドの特性は、質量分析に加えてまたはそれに代わって、IRまたはFTIRによって決定される。これらの技法により、リガンドの官能基または置換基（例えば、水酸基またはアミノ基）が同定できる。質量スペクトルと併用すると、同一分子量のリガンド間の相違を容易にさせ得る。

【0106】

本発明によれば、リガンド-標的対の解離定数 (K_d) は、約100 μM 未満、好ましくは、約10 μM 未満であるべきである。リガンド-標的対の解離定数 (K_d) は、標的分子機能の

50

決定、また薬物リード化合物として、リガンドを利用する方法に精通した人々を、リガンドの有用性を決定するのに際して、決定的ではないが、ガイドする因子の1つである。それ故、本発明では、その解離定数 (K_d) が約1 μ M 未満、約100 nM未満、約 10 nM未満、約1 nM未満、約 100 pM 未満、約10 pM未満のいずれかである、リガンドー標的対の相互作用を意図しているが、必ずしも選ぶ必要はない。

【0107】

適切な親和性を持つヒット化合物が無いが、少数しか見出されない場合でも、化学ライブラリの構造的多様性の中から構造的または化学的ギャップがすでに同定されていることもあり得る。そのような場合、標的分子の直接合成を利用して、そのギャップを埋めることができる。低い親和性の結合体が見出された場合、その結合を1つの機能的ドメイン上で、光活性化（または他の）リンカーを含むライブラリと共に繰り返すことができる。最初のカラム使用後に、タンパク質とそれに結合している分子のみが存在している場合、光活性化操作が実施でき、その後小分子が逆相HPLCで溶出できる。この方法で、標的は、テンプレートとしてすでに使用されており、低い親和性で結合している2つの分子にリンクしているもので、それらの分子は標的に対する親和性が増大している。好ましい実施例で、親和性は2倍～100倍増大する。

【0108】

5.1.4.1. 代表的な化学的アレイアッセイの実験方法と結果

HPLCに基づくアッセイ法

薬物様化学的骨格の収集物を表す薬物様化学的化合物 (Sigma-Aldrich, ICN, Calbiochem) の重さを量り、各最終濃度が20 μ M となるように50 mM 酢酸アンモニウムpH 7の10%メタノール溶液中で混合した。1 μ M～20 μ M チューブリンまたはP38 MAP キナーゼ (Sigma) をHPLC用 少量試料キュベットに入れ、0.5 μ M～20 μ M の化合物と混合した。混合し37℃で15分間インキュベーション後、キュベットを氷上に置き、自動注入装置 (Waters) でHPLC (Waters 2690) に注入した。HPLCの条件は、デュアルサイズ排除および相分離型150 mm X 2.1mm ID Pinkerton GFF II カラム (Regis Technologies) を使用し、使用バッファーは50 mM 酢酸アンモニウム10%メタノールであった。標的タンパク質と結合化合物はDiodeアレイ検出器によって検出され、カラムのボイド (空隙) 容量内で溶出され、ほとんどの化合物は波長243 nm でよく吸収された。ある場合には、低濃度の各化合物 (0.5～5 mM) および1個ずつの分離が容易と思われる10個未満の化合物を低濃度で使用すると、2つの標的タンパク質を滴定し、さらに標的タンパク質の1つに結合し、非特異的な対照物質には結合しないことが既知な特異的化合物のUV吸収波長においてその滴定を観察することが可能であった。

【0109】

本発明者らは、至適なカラムサイズおよび至適な樹脂を選択し、標的タンパク質と結合した化合物を非結合化合物から最大限に分離した。ボイド容量でタンパク質を溶出する樹脂、ボイド容量を最小限にする小径で短いカラムが使用された。このようなカラムによりタンパク質試料の希釈量を最小限にし、各アッセイの所要時間を最小にし、その結果、タンパク質から解離する結合化合物の量を最小限にする (K_{off} 定数に依存)。これらの特徴により、最小量の試薬と感度の高い検出法の使用が可能となる。タンパク質が2～3分未満で溶出するようなカラムの長さを使用した。Regis 150 mm x 2.1 mm GFF II カラム、1.0 mm x 100 mm YMC Diol カラム、2.1 mm x 150 mm Phenomenex ポリヒドロキシメタアクリレート (Polysep) カラム、およびJordi 2.1 x 150 mmジビニルベンゼンカラムを含む多くのHPLC カラムを試験した。同様に、塩とメタノール濃度を変化させ他のバッファーも試験し、結合反応に対する標的タンパク質と小分子の比率も1000:1～1:1000まで変化させた。異なるクラスの代表的な樹脂を使用して、薬物様小分子化合物からタンパク質画分を分離させる能力、さらに全化合物をカラムから溶出させるサイクル時間を最小にする能力も試験した。これらのカラムの特徴は、表面の特性と、背圧下で樹脂が破壊することによる流速の限界によって決定される。YMCジオールカラムは、シリカベースのカラムで、圧に対する抵抗性があり、サイクル時間が10分未満であるが、図9に挙げた100種化合物

の混合物のほんの約50%しかタンパク質から分離できなかった。Phenomonexポリヒドロキシメタアクリレートカラムは、100種類の化合物の約80%をタンパク質から分離でき、多数の小分子化合物の溶出を達成するには、メタノール勾配が必要であった。つまり、600 psiを超える背圧には耐久不能なため、比較的低い流速 (0.18 ml/min) で実施した。Phenomonexカラムのサイクル時間はメタノール勾配を使用すると1.5 時間であり、勾配を使用せず単離できる化合物のサブセット (全体の15%) に対しては35分であった。他のポリマー使用カラム (例えば、ポリヒドロキシメタアクリレート (Phenomonex, Shodex, Waters社)、ポリメチルメタアクリレート (Shodex, TosohBiosep)、Sephacrose/Sephadex/Supero se (Amersham Pharmacia Biotech)) も比較的低い流速でのみ耐久性があった。Jordi D VB カラムはジビニルベンゼンポリマーカラムであり、高圧 (4000 psi) で操作できるが、化合物同様にタンパク質も不要に結合してしまい、その結果、使用したバッファー系では分離ができなかった。非結合化合物からタンパク質を分離させるために他のバッファー系の使用が期待される。異なるカラムと樹脂が連続して組み合わせられ、化合物のタンパク質からの分離度合いを増加させたが、サイクル時間も長くなった。サイクル時間が長くても (例えば、1測定に10分を超える) 構わない場合は、上述のカラムまたは上述のカラムシリーズが使用され得る。

【0110】

短いサイクル時間が必要な場合、他のカラムが使用され得る。例えば、Regis GFF II カラムは、測定した化合物の97%からタンパク質画分を分離した。8000 psiの圧定格は、これらのアッセイに使用するHPLC (Waters 2690) の定格、つまり6000 psiで操作する、を超えていた。この樹脂のサイクル時間は、容易に8 分未満を示したので、圧力8000 psiまで耐久性のあるHPLCでは、より高速な流速を使用すると、さらに時間が短縮される可能性がある。GFF II樹脂および GFF樹脂は、Thomas Pinkerton が、タンパク質吸着による妨害なしで、血清中の薬物および薬物代謝物の直接分析を目的として開発した内部表面逆相樹脂である。この樹脂は、分子量12,000 ダルトン未満の分子のみが内部に接近できる、外部表面が親水性で内部ポアが疎水性な多孔質なシリカ支持体から構成されている。これらの表面は、グリシン-フェニルアラニン-フェニルアラニン (GFF)、またはグリシンドキシルプロピリン-フェニルアラニン-フェニルアラニン (GFF II) のトリペプチドがシリカ表面に結合することにより、生成される。GFFまたはGFF II を骨格とするビーズは、続いてエキソペプチダーゼであるカルボキシペプチダーゼAで処理される。この酵素は、ポアから排除されるほど十分大きな分子量 (35,000 ダルトン) を持ち、その結果、外表面からフェニルアラニン-フェニルアラニン部分が開裂する。この処理により、グリシンまたはグリシンドキシルプロピル基は外表面にそのまま曝露され、外面は親水性になり内面にはそのまま元のトリペプチドが残り、その結果、内表面は疎水性となる (例えば、製薬業者の添付文書に記載されている)。使用したGFF II樹脂カラムのカタログ番号は288-4である。これらの樹脂でバックされた他のカタログ番号のカラムもRegis technologiesで販売しており、使用できる。それで、外表面は、サイズ排除および親水性相互作用によって、大分子が内部相に入ることを防ぐ。疎水性作用に基づき化合物を保持し分離する、疎水性支持体を含む内表面に、小分子が入る。GFF II樹脂で実行できる短いサイクル時間と分離の程度から、GFF II カラムが次のアッセイに使用された。しかし、他の樹脂も使用できる。

【0111】

HPLC カラムより得られたタンパク質画分を1%TFAに解離させ、100uLの試料が逆相カラム (Waters Symmetry Shield) に注入され、タンパク質に結合していた化合物が分離された。化合物はアセトニトリル勾配によって、UV検出器を通過後TOF質量分析計 (Micromass LCT) に溶出された。バックグラウンドシグナルは、化合物が存在しないタンパク質を含む対照物質を使用して、各試料から差引かれた。質量スペクトルは化合物の分解が起こるのに十分なコーン型電極電圧 (20~80 ボルト) で測定された。他の質量分析装置でも、衝突セル中で分解は可能である。各化合物の特徴的分解パターンは、化学物質またはその同位体の分解物を示す、大きな親ピークと他のピークからなる。標的タンパク質から遊離さ

れた化合物の分解パターンは、標準物質で観察された分解パターンの特徴と比較され、標的タンパク質に結合している化合物が同定された。また、化合物の分子量を示す親ピークに対する1つ以上の同位体の特徴的ピークが標準物質と比較され、標的タンパク質に結合する化合物が同定された。代替の別の分析では、化合物の分子量を示す親ピーク自身を標準物質と比較し、化合物を同定した。時々、これらの方法を併用して化合物が同定された。化合物の分解を誘発しないMS条件下で類似な方法が適用され、その結果、その化合物（例えば親ピーク）の分子量とその同位体を示すピークを含む質量スペクトルが得られた。

【0112】

HPLCに基づいた方法による測定結果

SKB86002は、P38 MAPキナーゼ標的タンパク質とマイクロモル単位で親和性のあるリガンドである。P38 MAP キナーゼ (5 μ M) を、5 μ Mの 86002 と混合させ、ジオールカラムのHPLCで分離した (図3)。タンパク質画分を採取し、質量分析計で分析した。スペクトル上の親ピーク、フラグメント、および同位体ピークは86002の標準物質に相当し、このことは、P38 MAP キナーゼがマイクロモル単位の親和性を持つ特異的なリガンドを単離し、抽出していることを示している。

【0113】

SKB86002とキニン1塩酸塩（非特異的な対照物質）を各々の最終濃度が5 μ M となるように混合した (図4)。P38 MAP キナーゼタンパク質を増量させながら（最終濃度0、2.5、5 および 10 μ M）混合化合物と混合し、各最終濃度を5 μ M とし、タンパク質をジオールカラムHPLC で分離した。UVスペクトル上で、86002ピークはP38濃度依存性の減少を示したが、キニンのピーク減少は無視できるほどであった。

【0114】

P38 タンパク質画分を、図4で示す滴定（5 μ M のP38 MAP キナーゼ + 5 μ Mの キニンと86002の混合物）の midpoint で採取する場合、混合物から抽出され、タンパク質から遊離された化合物は、遊離化合物の質量ベクトル上の親ピーク、フラグメント、および同位体ピークに基づいて、キニンではなく86002と同定された (図5)。

【0115】

86002とコルヒチンを含む10種の薬物様化合物の等量混合物を調製した (図6)。P38 MAP キナーゼタンパク質を増量させながら（最終濃度0、3.5、および5 μ M）、この10種の混合化合物と混合し、各々の最終化合物濃度を0.5 μ M とし、タンパク質をGFF II カラムHPLC で分離した (図7)。UVスペクトル上で、86002ピークはP38濃度依存性の減少を示したが、コルヒチンのピークまたは混合物の他の化合物を表すピークの減少は無視できるほどであった。タンパク質画分を採取し、質量スペクトルを測定すると、そのスペクトルには、他のピークよりもはるかに高い強度に86002の特徴を示す親ピークと同位体ピークが現れた。

【0116】

チューブリンタンパク質を増量しながら（最終濃度0、5、20 μ M）この10種の混合化合物と混合し、各々最終化合物濃度を0.5 μ M とし、タンパク質をGFF II カラムHPLC で分離した (図8)。UVスペクトル上で、コルヒチンピークはチューブリン濃度依存性の減少を示したが、86002 のピークまたは混合物の他の化合物を表すピークの減少は無視できるほどであった。タンパク質画分を採取し、質量スペクトルを測定すると、そのスペクトルには、他のピークよりもはるかに高い強度にコルヒチンの特徴を示すピークが現れた。

【0117】

86002 とコルヒチンを含む100種の薬物様化合物の等量混合物を調製した (図9)。P38 (2 μ M) を100種の混合物と混合し、各化合物の最終濃度を20 μ M とし、GFF II カラムのHPLCによって、タンパク質を未結合化合物から分離した (図10)。タンパク質画分を採取し、化合物をタンパク質から遊離し、質量スペクトルを測定した。スペクトルには、他のピークよりもはるかに高い強度に86002の特徴を示すピーク得られた。このことから、P38 MAP キナーゼは、特異的な濃度依存性様式で、100種の混合物からマイクロモル単位の親和性

10

20

30

40

50

を持つ1つのリガンド (86002) と結合し、抽出する。質量スペクトルのバックグラウンドは、10種混合物 (図7) が示したスペクトルの場合に匹敵するように見え、このことは、アッセイがより多数の化合物にスケールアップできる可能性を示している (例えば数千から数万種類の化合物)。例えば、これらの方法は、10、20、40、50、75、100、200、500、1000、2000、5000、10000個、もしくはそれ以上の化合物のライブラリ、またはより多くの化学骨格構造のライブラリを分析するのに使用でき得る。

【0118】

チューブリン (5 μ M) を100種類の化合物混合物と混合し、各化合物の最終濃度を5 μ M とし、GFF II カラムのHPLCによって、タンパク質を非結合化合物から分離した (図11)。タンパク質画分を採取し、化合物をタンパク質から遊離し、質量スペクトルを測定した。スペクトルには、他のピークよりはるかに高いレベルに、コルヒチンの特徴を持つピークが現れた。このことから、チューブリンは、特異的な濃度依存性様式で、100種の混合物からヒット化合物 (コルヒチン) と結合し、抽出する。質量スペクトルのバックグラウンドは、10種混合物が示したスペクトルの場合に匹敵するように見え (図8)、このことは、アッセイがより多数の化合物にスケールアップできる可能性を示している (例えば数千から数万種類の化合物) 例えば、これらの方法は、10、20、40、50、75、100、200、500、1000、2000、5000、10000個、あるいは それ以上の化合物のライブラリまた、より多くの化学骨格構造のライブラリを分析するのに使用でき得る。

【0119】

アッセイの速度を上げる方法は、流速を高くすることである (図12)。カラムが耐えられる最大流速に影響する限定要因は、一般的に樹脂が破壊する前に耐えられる背圧である。GFF II樹脂を選択する理由の1つは、最大背圧が100~1500 psiであるサイズ排除ゲル (例えば、Sephacrose, Superose, Superdex、ポリメチルメタアクリレート、ポリヒドロキシメタアクリレートなど) と比べて、最大8000 psiまで耐久できるからである。GFF II カラムは高流速であっても、100種類の化合物からタンパク質は良好に分離された。

【0120】

スピнкаラムクロマトグラフィー

薬物様化学骨格構造 (Sigma-Aldrich, ICN, Calbiochem) の収集を表す薬物様化学物質の重さを計り、各物質の最終濃度が20 μ M となるように、pH 7の50mM酢酸アンモニウム10%メタノール溶液に混合した。5 μ M~20 μ M ウシ血清アルブミン (BSA) またはチューブリン (Sigma) をHPLC用 少量試料キュベット (Waters) に入れ、5 μ M~20 μ M の化合物と混合した。混合後、37°Cインキュベーションを15分間し、キュベットを氷上に置いた。図9に挙げた100種混合化合物の50 μ L を、あらかじめ結合バッファーで2回洗浄し (すなわち、各洗浄操作では、50 mM 酢酸アンモニウム10%メタノールバッファー200 μ L を加え、さらにカラムを通過させるようにしたバッファーを1.5 mL の微量遠心管 (Eppendorf) 中で、回転数の最大設定値において30秒から1分間微量遠心機 (Eppendorf) で遠心させる)、バッファーと平衡状態にしてある、MicroSpin G-25 (Amersham Pharmacia Biotech) スピнкаラムの上清に添加された。このようなスピнкаラムは、一般的に標識化後DNAプローブ用にバッファーの脱塩および交換に使用されるが、G-25 は分子量が25KD以上の分子を削除する従来のサイズ排除樹脂の1つである。スピнкаラムを続いて1.5 mL の微量遠心管 (Eppendorf) に入れ、微量遠心機 (Eppendorf) の最大設定で30秒間遠心させた。また、吸引させるとスピнкаラムから溶液を引き出すことができ、スピнкаラムは、スピнкаラム/カートリッジを96 (個) ウェル型で整列し、吸引マニホールドを使用して、溶液をカラムから96 (個) ウェルプレートに引き出す場合に特に有用である。

【0121】

BSAを用いる場合、微量遠心管の底部にある50 μ L溶液をHPLCに注入し、UVスペクトルを可視化して分離前のBSA/100 化合物混合物に相当する量と比較した。チューブリンを用いる場合、微量遠心管の底部にある25 μ L の溶液を1%TFA で解離させ、逆相カラム (Waters Symmetry Shield) に注入し、化合物をアセトニトリル勾配によってUV検出器を通過させ TO F MS (Micromass LCT) に溶出させた。バックグラウンドシグナルは、化合物が存在しな

タンパク質を含む対照を使用して、各試料から電子工学的に差引かれた。質量スペクトルは化合物の分解が起こるのに十分なコーン型電極電圧（20～80 ボルト）で測定された。他の質量分析装置でも、そのような分解は衝突セル中で可能である。各化合物の特徴的分解パターンは、化学物質またはその同位体の分解を示す、大きな親ピークと他のピークからなる。標的タンパク質から遊離された化合物の分解パターンを、標準物質で観察された分解パターンの特徴と比較し、標的タンパク質に結合した化合物が同定された。また、化合物の分子量を示す親ピークに対する1つの特徴的同位体を標準物質と比較し、標的タンパク質に結合した化合物が同定された。代替の別の分析では、化合物の分子量を示す親ピーク自身を標準物質と比較し、化合物を同定した。時々、これらの方法を併用して化合物を同定した。化合物の分解を誘発しないMS条件下で類似な方法を適用し、その結果、その化合物（例えば親ピーク）の分子量とその同位体を示すピークを含む質量スペクトルが得られた。

10

【0122】

スピнкаラムクロマトグラフィーに基づいた方法の結果

ウシ血清アルブミン（BSA, Sigma）5 μ M を、各化合物の最終濃度が5 μ M となるように100種化合物に混合した（図13）。混合物の半量（50 μ L）をMicro-Spin G-25 カラムの上部に層状に添加し、遠心分離した。タンパク質を含む画分を遠心分離管の底部に採取した。初回のタンパク質/化合物混合物をスピнкаラム分離法によって分離した後のタンパク質/化合物混合物と比較し、タンパク質が十分生成されていることをUV吸収から観察した。同じプロトコールを20 μ M のチューブリンと20 μ M の100種混合物との混合に適用し、溶出タンパク質含有画分の質量スペクトルを測定すると、コルヒチンの特徴を持つピークが、他のピークよりはるかに高い強度に生じた。この場合のバックグラウンドピークは、HPLC カラムの分離によって観察されたものよりも僅かに高かったが（図14）、スピнкаラムによる分離速度および測定の変動性は非常に魅力的である。例えば、これらの方法は、10、20、40、50、75、100、200、500、1000、2000、5000、10000個、あるいはそれ以上の化合物のライブラリまた、より多くの化学骨格構造のライブラリを分析するのに使用でき得る。

20

【0123】

5.1.4.2. 単離リガンドを同定するパターン認識ソフトウェアの代表的な使用方法

本発明は、標的タンパク質と本書で記載した分離方法を用いて単離した混合物から、1つの化合物を同定するための、質量スペクトルのパターン認識分析の方法を提供する。

30

【0124】

これらの方法では、候補化合物の最初の混合物中に存在する多数またはすべての化合物について、質量スペクトルの分解パターンが測定される。また、同位体あるいは他の質量スペクトルパターンもこれらの化合物について測定される（例えば、M+1またはM+2 同位体ピーク）。質量分析計は、化合物、その同位体、および/またはそのフラグメントを、 m/z で表す質量/電荷比に基づいて分類する。ほとんどあるいはすべてのピークが+1（または-1）の電荷を持つ分子となるように、質量分析の条件を調整できるので、その結果、いくつかのピーク値は親化合物、同位体または親化合物のフラグメントの質量に等しくなる（すなわち、 $m/z = m/1 = m$ ）。ある場合には、いくつかの、またはすべてのピークが+2 以上（または-2 以下）の電荷を持つ分子を示すように他の質量分析の条件を使用し、その結果、質量/電荷比が分子の質量未満となるため（例えば、 $m/z = m/2$ ）、いくつかのピークの値は親化合物、同位体、またはフラグメントの質量より小さくなる。そのため、質量分析パターンは、親化合物、そのフラグメントおよび/またはその同位体の質量（または分子の電荷が1を超える場合は質量/電荷比）に相当する質量スペクトルピークから構成される。

40

【0125】

これらの各ピークの質量（または質量/電荷比）を情報検索システムのデータベースに入力する。標的タンパク質から遊離した対象化合物の質量スペクトルが得られ、続いてパターン認識ソフトウェアを使用して、測定したパターンとデータベースのパターンを比較す

50

る。明確にマッチすると、対象化合物が確実に同定される。1つの実施例では、2個、3個、またはそれ以上の最も特徴的な質量に相当するピーク（化合物1に相当するピークA、B、C、また化合物2に相当するピークD、Eなど）を、最初の混合物中の各化合物のデータベースとして入力する。ソフトウェア（例えば、Micromass社のMassLynxバージョン3.5）を使用して、標的タンパク質から遊離した化合物の質量スペクトルのピークAを検索し、続いてピークB、C、D、Eを順番に検索する。特定ピークの存在を第二のデータベースに入力し、そのピークが質量スペクトル上に現れていることを示す。別の可能な方法では、質量スペクトルの特定ピークの検索は、どんな順序でも実施される。質量スペクトルの解析には、対話型検索コマンドが使用され得る。例えば、特定化合物に相当するピークAが質量スペクトル上に生ずる場合、そのスペクトルを解析し、同じ化合物に相当する別のピーク（例えばピークB）の特徴がスペクトル上に出ているかを確認し得る。また、特定化合物に相当するピークの特徴が質量スペクトルに出ていない場合、そのスペクトルを解析し、別の化合物のピーク（例えばピークD）の特徴がスペクトル上に出ているかを確認し得る。さらに別に代わる方法では、マクロプログラムをMassLynxにオーバーレイして、複数のピークを同時に検索している。現時点で同定されたピークを、最初の混合物中の化合物より得られた最初のデータベース内のピークと比較し、標的タンパク質から遊離される化合物を同定する。図16 Aに、これらの方法の数例の実施例に対し、その操作を図解した代表的なフローチャートが記載されている。

【0126】

別の態様では、質量分析パターンの最も特徴的なピークに相当する2個、3個、またはそれ以上の質量数値（または質量/電荷比）が、最初の混合物中の各化合物のデータベースとして入力される。代表的な方法において、このデータベースはMicrosoft ExcelまたはOracleのプログラムを使用している。一旦、標的タンパク質から遊離された試料の質量スペクトルが測定され、そのスペクトル上に2つまたは3つのピーク（例えば、最も高いシグナルを持つ2つまたは3つのピーク）の位置が確認されると、これらのピークに相当する質量数（または質量/電荷比）を用いて、最初の混合物のデータベースから検索が実行される。例えば、質量数を、プログラムの「検索」コマンドで入力し、その質量のピークを生ずる候補化合物の検索ができる。その検索で同定された質量数の組み合わせから、試料中に存在する化合物を同定する。

【0127】

さらに別の態様では、特定の質量（または質量/電荷比）のシグナル強度を使用して、化合物が正しく同定されている。この技法は特に、使用パターンが同位体の場合に適用される。この場合、混合物中の化合物のデータベースが生成され、2つまたは3つの特徴的なピークの各質量と強度の両方を含んでいる。対象試料に関するこれらの情報が収集される。データベースプログラムの検索機能を利用して、質量と強度の相関パラメータが検索される。明確にマッチすると試料中に存在する化合物が確実に同定できる。

【0128】

対象となる1つ以上の化合物（例えば、1つの標的から遊離された化合物）の同定に本発明方法を使用した様々な態様において、1つの化合物の1つ以上のフラグメントに相当する1つ以上質量スペクトルピーク、および/または1つの化合物の1つ以上の同位体に相当する1つ以上の質量スペクトルピークを利用してその化合物を同定している。他の実施例では、親ピークが化合物の同定に使用されている。様々な実施例で、親ピークは化合物の同定に使用される唯一のスペクトルピークである。さらに他の実施例で、化合物の同定において、親ピークは、フラグメントまたは同位体に相当する1つ以上のピークと併せて使用される。さらに他の実施例では、親ピークが化合物の同定に使用されていない。他の実施例では、化合物は対象標的と接触した少なくとも5、10、20、40、50、75、100、200、500、1000、2000、5000、10000個、あるいはそれ以上の化合物との混合から回収された構成成分である。他の実施例では、化合物は少なくとも、5、10、20、40、50、75、100、200、500、1000、2000、5000、10000、あるいはそれ以上の異なる化学骨格構造を含む化合物の混合から回収された構成成分である。特殊な実施例では、親ピークは、少なくとも5、10、2

0、40、50、75、100、200、500、1000、2000、5000、10000、あるいはそれ以上の異なる化学骨格構造を含む化合物の混合から、1つの化合物を同定するために使用される。

【0129】

ここに記載されている方法はすべて、実際、いずれのコンピュータによっても実行され得る。図15は代表的なコンピュータシステムを示している。コンピュータシステム2は内部および外部コンポーネントを含んでいる。内部コンポーネントは、メモリ6に繋がったプロセッサ4である。外部コンポーネントは、質量データ保存装置8、例えば、ハードドライブなど、ユーザ入力装置10、例えばキーボードやマウス、ディスプレイ12、例えばモニター、および通常、コンピュータと他のコンピュータを繋ぎ、データや処理作業を共有することができる、ネットワークリンク14である。プログラムは、運転中本システム2のメモリ6に搭載される。これらのプログラムには、オペレーティングシステム16、例えばコンピュータを管理するMicrosoft Windowsなど、本発明の方法を実行するプログラムを補助する共通言語および機能をコードするソフトウェア18、および操作言語または記号パッケージで本発明の方法をコードするソフトウェア20が含まれる。それらの方法をプログラムするために使用できる言語は、Microsoft社のVisual C/C++ など、特に限定されない。好ましい適用では、本発明の方法は、プログラムの実行に使用されるアルゴリズムを含んで、式の記号入力および特異性の高い処理が出来るような数学的ソフトウェアパッケージによってプログラム化され、その結果、ユーザーが、個々の式またはアルゴリズムの手順をプログラムする必要性がなくなる。この目的に役立つ代表的な数学的ソフトウェアパッケージはMathworks社(Natick, MA)のMatlabである。Matlabによって、Parallel Virtual Machine (PVM) モジュールおよびMessage Passing Interface (MPI) を適用でき、複数のプロセッサでの処理をサポートできる。本書の方法と共にPVMおよびMPIを実行するには、既存の方法を使用する。また、ソフトウェアやその一部が、既存の技法による専用回路でコードされる。

【0130】

5.1.5. 標的分子機能の解析

標的分子機能を系統的に分類するために、各標的のヒット化合物を細胞や組織を使用したアッセイによってスクリーニングし、疾病病因に対する主要分子の各メカニズムを示し得る。標的が示差的発現解析法に基づいて最初に選択される場所で、その示差的発現に特に関連するアッセイが好ましい(例えば、がん細胞の示差的発現解析から標的が生じた場所に特に関連することもある)。この一連のアッセイ法には、アポトーシス、増殖、虚血/壊死、炎症、線維症、血管形成、代謝シグナル伝達、感染および発生/分化などを検出または測定するアッセイを含むが、これだけに限定されない。病因発生経路に注目し、疾病および細胞に特異的な標的を研究することにより、多数の治療領域に対する新しい標的分子が同定され得る。このパネルの目的は、重要な疾病、つまり慢性変性疾患(例えば、アルツハイマー病、変形性関節症、骨粗鬆症)、代謝疾患(例えば、糖尿病、肥満)、炎症疾患、がん、心血管系疾患(例えば、冠動脈疾患、高血圧、うっ血性心不全、心筋症、慢性腎不全)および感染症(例えば、ウイルス、細菌、原虫および薬剤耐性のメカニズム)を含み、これだけに限定されないが、これらの原因となる分子経路の小分子/タンパク質をスクリーニングすることである。アッセイは、同じアッセイが最初に細胞で使用され、疾患患者の生検組織での追跡検査が行えるようにデザインされる。有毒な可能性のある分子を同定するために、壊死アッセイがすべての分子に実施され得る。業界で標準の96ウェルのマイクロタイタープレートは、ハイスループットおよび超ハイスループットも除外されないが、表現型スクリーニングを実施するために十分なスケールを提供している。アッセイは細胞株、一次細胞培養、組織生検、組織モデル、インビボ動物モデル、または他の生物で実施され得る。好ましい実施例では、生物学的アッセイは、ヒト細胞株と組織を使って実施される。他の実施例によれば、生物学的アッセイは、細胞、組織、臓器またはいかなる種族の生物全体を使って実施され得る。これらのアッセイでリガンドがプールできるが、各表現型のアッセイは、1つのウェルに付き1種の分子に対し実施され、表現型の作用を隠すようなアゴニストならびにアンタゴニスト相互作用を避けることが有用であ

る。アッセイでは、疾病細胞または組織を、疾病や治療反応に関連すると思われる遺伝子で豊富にさせるが、これだけに限定されない。

【0131】

がん、糖尿病、およびTGF β による細胞刺激などに対する標的分子同定への、本発明の適用が実施例に記載されているが、上述のアプローチはいずれの疾患、細胞刺激、生体調節物質（モジュレータ）または病態に対しても広く適用することができる。上述以外のアッセイまた、疾病に関連する他の分子経路のアッセイも使用できる。正常細胞や組織と相対的な疾病細胞、またはアゴニストやアンタゴニスト（またはそれらの一部）の存在する細胞内で、上方制御または下方制御される遺伝子を使ってこのアプローチを開始すると、特異性や良好な治療指数を持つ標的が豊富になる。この特異性を疾病病因の分子メカニズムと考え合わせると、治療に有用な標的が豊富となる。大きなライブラリから効率よい方法でヒット化合物を選択する、生化学的結合アッセイを順番に組み合わせ、またヒト疾病を反映する低スループット高品質表現型生物学的アッセイにおいてこれらのヒット化合物を使用すると、遺伝子機能を決定できる。

【0132】

5.2. 表現型から遺伝子型へ

別の一連の実施例において、本発明は、少なくとも1つの生物学的アッセイで多数の可能性のあるガンドをスクリーニングし、1つの生物学的アッセイで表現型の変化を生ずるリガンドを選択し、さらにそのリガンドを使用して標的候補物質をスクリーニングし、変化した表現型の原因となる特定の標的物質を同定する方法に関連する。様々な好ましい実施例で、個々のリガンド種は生物学的アッセイで別々にスクリーニングされる。生物学的アッセイで表現型に変化を与えるリガンドが、リガンド-標的相互作用を起こさせる条件下で多数の可能性のある標的に曝露され得る。本発明の様々な好ましい実施例において、標的はペプチドまたはタンパク質であり、各ペプチドや標的タンパク質はその標的をコードするポリヌクレオチドに関与している（例えば、ファージディスプレイまたは細胞表面ディスプレイ）。選択した標的とそれに対応するポリヌクレオチドが採取される。タンパク質の標的をコードするDNA配列が決定され、クローン化され、確認される。これら標的の示差的発現は、ヒト疾病組織生検、特に表現型の分子メカニズムが表現型に関連しているような組織で試験され得る。同様に、リガンドも、これらの疾病組織および/またはインビトロまたはインビトロの疾病モデルにおいて試験され得る。1つの実施例の概要を図2に示す。上述したように、5.1.1～5.1.5の項で挙げた実施例はこれらのいずれの方法でも使用できる。

【0133】

本発明によるハイスループット表現型細胞に基づくアッセイは、現在行われているハイスループットスクリーニング法と異なる。典型的なハイスループットスクリーニング法は、確認した標的の遺伝子がリポーターシステム（例えば、緑色蛍光タンパク質、ルシフェラーゼなど）によって細胞株に形質移入され、そのリポーターの活性から化学ライブラリのメンバーがスクリーニングされるという、メカニズムに基づくアッセイである。このタイプのスクリーニングを実施する代わりに、本発明では、1つの生物学的アッセイによる分子標的の予備的測定をせずに、細胞株における表現型の有意な変化を探すことに着目している。これらの生物学的アッセイは、重要な生体刺激物質または重要な発病メカニズムを調節するリガンドを探すことを目的としてデザインされている。限定のない実施例には、アポトーシス、増殖、虚血、壊死、炎症、線維症、浸潤、血管形成、代謝、感染および胚形成などが含まれる。さらに、多分化能効果を持つ細胞刺激物の個々の経路は、アンチセンス、転移ペプチド、抗体またはそれらの作用により特異的な標的を同定する他の技法によって、遮断することができる。この方法で本発明者らは、1つの生物学的アッセイからライブラリ（上述したような）のリガンドと表現型との関連性を得る。上記事項など、ただしそれだけに限定されないが、疾病の分子レベルメカニズムのアッセイは、ハイスループットスクリーニングに採用され得る。

【0134】

10

20

30

40

50

癌の標的同定に、本発明を適用することが本書で検討されているが、本発明はいかなる疾病、細胞刺激または病態に広く適用できる。生体刺激に関連した上述以外の他のアッセイ、さらに疾病または生物学に関連する他の分子経路のアッセイも使用できる。リガンドが対象とする特定の表現型変化に関与している生物学的アッセイを順番に組み合わせ、さらにこれらのヒット化合物を使って、タンパク質またはペプチドのディスプレイライブラリから標的分子を選択すると、標的の遺伝子をクローン化し同定できる。ヒト疾病組織における標的の示差的発現が、引き続き試験され得る。さらに、インビトロまたはインビボの生物学的アッセイにおけるリガンドの特異的作用から、生体への影響の調節あるいは特定疾患の治療における、そのリガンドの有用性が明らかにされ得る。

【0135】

5.3. シグナル経路分子のマッピング

一旦、多くの遺伝子が発病の特定な分子経路に関連することが示されると、その標的分子は、相互に、さらに経路の既知メンバーに関連して分子経路内でマッピングされる。異なるタンパク質と結合するリガンドは、光活性化架橋剤で誘導体化され、経路中の各メンバーを位置づけるのに使用され得る。例えば、経路の1つのメンバーは最初に標識化される（例えば、GFP）。次に、架橋が可能な官能基を持つ誘導体化リガンドに、経路のメンバーが曝露される。続いて、その混合物が架橋刺激物に曝露される。最後に、経路内の選択されたメンバーが、標識化（例えば、GFP）によって採取され、それに会合するようになるなどの化合物も同定される。この段階が、前部のまたは後部の経路メンバーを同定するために、順番に繰り返され得る。これらの方法は、架橋に先立って、前もってリガンドの結合部位を同定したり、標的分子の二次または三次構造を決定する必要がない利点を有する。

【0136】

経路メンバーは続いて、リガンドスクリーニングにおいて標的として使用され得る。各経路メンバーに選択的に結合する、各リガンドの表現型を比較することによって、他のメンバーと相対的な各メンバーの位置情報が得られる。この情報を利用して、特定の疾患適応症に対する最良の標的分子を確認・選択し、最終的に、薬理遺伝学に基づく診断を通して最適な治療法を選択し得る。

【0137】

5.4. リード化合物の最適化

本発明は、リード化合物を最適化し、ヒットする割合を上げる方法を提供する。ここで、「リード化合物」とは、製薬的に好ましい特性を持つリガンドを意味する。好ましくは、そのリガンド分子は、技術的に「小さい」分子、例えば分子量が50～3000 ダルトンの間の分子と考えられる。この方法は広い応用範囲を持っているが、タンパク質-タンパク質相互作用を妨害するリガンドを得るのに、特に有用である。

【0138】

多数のリード化学物質が生化学的および表現型レベルで特徴付けられるので、構造活性相関が確認され、リード化合物の最適化の基礎となり得る。類似活性を持つ分子が同定されると、構造活性相関（SAR）が決定できる。標的を目的とした合成技術を使用して、お互いに近くで結合している分子を架橋することができ、その結果、結合分子の活性が、タンパク質の同じサブサイトまたは、標的タンパク質の異なるサブサイトを通して仲介されるか否かが示される。1つの実施例で、分子の1つは光活性化架橋剤を含むかまたは、2番目の分子上の基と反応する反応基を含んでいる。この方法で、標的上で追加される別の機能を持つサブサイトがマップされ、異なるメカニズムを、これらのサブサイトで結合する分子による表現型の結果から解釈することができる（例えばアゴニスト対アンタゴニスト）。リガンド構造骨格上の1つの官能基にある光活性化架橋剤を使用して、標的と結合しているリガンドをリンクさせ、その結果標的分子をテンプレート（鋳型）として使用し得る。

【0139】

この段階では、小分子A と小分子Bだけを混合することも、あるいは、標的と結合しない

10

20

30

40

50

他の小分子の存在下で、A、B両分子を混合することができる。この条件下では、1つの官能基が保護され他は無保護であり、A、Bのどちらにも反応する二官能性基架橋剤が存在する。また、Aは、架橋剤と反応し、その生成物はBと反応する可能性がある。官能基として、アミン、カルボン酸、ニトリル、およびハロゲンがあるが、これだけに限定されず、あらゆる反応基が挙げられる。A、Bには同じまたは異なる官能基が存在してもよい。互いに反応する小分子AとBとの1ペアについての1実施例において、Aはアミンの官能基を含み、Bはカルボン酸、活性化エステルおよび無水物、アシルハロゲン化物、またはアシル化またはアルキル化反応でアミドと反応する他の基を有する架橋剤を含む。リンカーは、2つの官能基のみを含むか、またはその官能基の間に1つの構成成分、ポリエチレングリコールなどを含むが、これに限定されない。代表的な保護基は、BOC、FMOC、またはベンジルのようなアミン保護基を含む。CBZ保護基を使用して、カルボン酸、ベンジルエステル、アリルエステル、およびニトリルを保護することができる。1つの実施例で、保護基は光活性化され、ニトロベンジルまたはアゾ基のような官能基を脱保護する。別の実施例では、タンパク質と反応しない官能基を含むリンカーと、タンパク質上に官能基を含まない化合物（例えば、アミン類、カルボン酸類、アルコール、SH基など）が使用される。1つの実施例では、化合物はハロゲンを1つ（例えばCl）含むか、またはハロゲンを1つ含むように修飾される。二重結合、三重結合、ハロゲン、または芳香族基を含むリンカーは続いて、Heckカップリング反応またはSuzuki反応を通じてその化合物にリンクされ、その結果、タンパク質と反応せずにリンカーと化合物の結合が起きる。このような化合物はAldrich社から販売されている。リンカーおよび上述の反応の保護基はAdvanced ChemtechやNovobiochem、その他から販売されている。この結合は、好ましい実施例では、標的への結合親和性を2~100倍以上増大させ得る。そのため、高い親和性の優れたリード化合物が得られる。このアプローチを使用して、標的に着目した生体に関連する方法で、化学ライブラリの構造的多様性をさらに高めることができる。

【0140】

6. 遺伝子型から表現型へ

6.1. 実施例1：乳がん

6.1.1. 標的物質

少なくとも1名の乳癌患者からまず生検を採取する。レーザー捕獲顕微解剖および、ANRNAまたはRT PCRをマイクロアレイ分析法と併用して、がん性細胞内で識別的に発現される遺伝子を単離し得る。例えば、これらの技法を使用して、同じ生検中に、非がん性細胞よりも2倍以上のレベルでがん細胞に存在する転写物を同定し得る。また、その遺伝子は非がん性細胞内では過剰発現され得る。また遺伝子として、試験患者の有意な画分からそのようなレベルで発現される遺伝子が選択され得る。

【0141】

組織はTissue Tek OCT 保存液（VWR）に埋め込まれ、液体窒素で凍結し、低温槽で切断され得る。切片は、非コーティングガラススライド上で固定し、-80℃で保存され得る。スライドは70%エタノール中で30秒間固定し、H&Eで染色し、続いて70%、95%、および100%キシレン中で5秒間脱水し、さらにキシレンで5分間脱水し得る。空気乾燥後、切片はPixCell I および II LCM システム（Arcturus Engineering）でレーザー捕獲顕微解剖され得る。形態学的に正常な乳房上皮細胞、悪性浸潤性乳がん細胞および悪性転移乳がん細胞（例えば、腋窩リンパ節）の各5 X 10⁴ が捕獲され得る。室温で接着細胞の付いた転移フィルムをイソチアン酸グアニジンに移動させ、フェノール/クロロホルム/イソアミルアルコールで抽出し、イソプロパノール中で酢酸ナトリウムと10 μg/μL グリコーゲンによって沈殿させることによって、総RNAが、これらの細胞の各集団から単離され得る。続いてRNAペレットは、再懸濁され、RNASE 阻害物質（Life Technologies）の存在下10ユニットのDNase（Gene Hunter）で37℃で2時間処理され得る。再抽出と沈殿後、ペレットは27 μLのRNASE不含の水で再懸濁され得る。ANRNAまたはRT PCRが実施され、続いて配列決定され得る。ESTであるこの技法によって同定された配列を使用して、cDNAライブラリ（CLONTECH）からcDNAの全長を選択し得る。これらのcDNAは、罹患した正常でない細胞/組織

に豊富に見られるが、それらの機能は未知であり得る。

【0142】

選択されたcDNAは、カルボキシ末端に挿入されたヘキサヒスチジン (6his)、および遺伝子のアミノ末端のグルタチオンシンセターゼ (GST) (各々にプロテアーゼ開裂部位を持つ) によってタグが付けられる。これらの遺伝子は、bipタンパク質リーダーによってショウジョウバエ発現系ベクター内でクローン化され、ハイグロマイシンベクターと共にCaPO₄によってショウジョウバエベクター内に共に形質移入され得る。細胞を選択培地に維持し、硫酸銅 (Invitrogen) を用いて遺伝子発現を誘導する。48時間後、5-10 mg/lの各タンパク質を含む上清を収集する。その後、得られたタンパク質を最初の精製段階としてその上清よりNi(2+)-NTA クロマトグラフィーにより精製し、二段目の精製法としてグルタチオンアフィニティークロマトグラフィーを使用し、その後タグ開裂による特異的なプロテアーゼの除去が行われる。最高ミリグラム量の各タンパク質が回収される。

【0143】

6.1.2. 結合、リガンドー標的対の選択、およびリガンドの同定

最大2百万個のリガンドを含む、多様な化学物質、天然物様物質およびペプチドの組み合わせライブラリが、液相中でプールされた様式で合成され得る。さらに、天然物ライブラリ (Terragen, Yonsei) および化学物質ライブラリ (Arqule, Coelocath) が購入され得る。1,000~10,000個のリガンドが最大100 μ l 中で1 μ g のタンパク質と共に混合され、96ウェルプレートの1つのウェル内で1 μ M の濃度となり得る。氷上で30分間インキュベーション後、試料は、カートリッジ付き96ウェルプレートに装填され、各ウェルにHPLCカラム (Waters 2790 HPLC) として操作され得る。最初のカートリッジ/カラムは、サイズ排除樹脂 (G25, Pharmacia) であり、樹脂中の非結合分子を保持するが、結合リガンドとタンパク質を通過させ得る。小型で狭いカラム (例えば、長さ2 mm x直径5 mm のRocke t Column, Biorad) を使用し、この操作の希釈を最小に留める。次に使用するカートリッジ/カラムは、疎水性または親水性逆相HPLC樹脂であり、その選択は、使用するリガンドライブラリの疎水性に依存する。例えば、疎水性C18シリカカラムは、疎水性が低いリガンドに使用され、親水性C8カラムは、より親水性の高いリガンドに使用され得る。別例として、疎水性または親水性どちらのリガンドにも使用され得る、Agilant 社のSB8Uカラムが挙げられる。逆層HPLCは、小分子とタンパク質を、樹脂に結合させることによりそれらを濃縮し、その後、小分子がタンパク質と樹脂から溶出され得る。小分子を含む溶出物は96ウェルプレートで採取され得る。これらの溶出物は、続いて、質量分析計 (Micromass Quattro LC) に移動され、MassLynx、MAXENT ソフトウェア (Micromass) によってスペクトルが測定され得る。この方法は理論的にいって、1つのタンパク質につき最大100個のリガンドが逆重畳積分され、正確なライブラリのメンバーそのものが、鏡像異性体を除き同定され得る。特に、質量分析は化合物の同位体または分解パターンを検出するために使用でき、その同位体・分解パターンいずれも代替え法として、または真の分子量と組み合わせ使用し、化合物を同定できる。さらに、IRまたはFTIR 分析を実施し、リガンドの官能基またはユニットが同定され得る。各リガンドが、続いて、合成されるか、または大規模スケールで合成され得る。ペプチドリガンドは、TAT形質移入配列で融合され得る。

【0144】

同定されたリガンドの親和性はスクリーニングで使用するライブラリの濃度に一部に依存するが、少なくともナノモルまたはマイクロモル単位の範囲であるはずである。各リガンドの実際の親和性は競合試験で決定され得る。これらのリガンドは続いて、生物学的アッセイで試験され得る。

【0145】

6.1.3. 生物学的アッセイ

cDNAががん細胞の示差的発現に基づいて選択される場合、リガンドは、アポトーシス、増殖、壊死、血管形成、炎症、または転移がん浸潤を検出・測定するアッセイで試験され得る。本発明によれば、アッセイは、ヒト疾病にできるだけ近く (例えば、病理組織の生検、インビトロ組織モデル、インビトロ疾病モデル、ヒト細胞株)、細胞株に基づきヒト

病理試料の一次組織に容易に適用されるモデルからデザインされる。これらのアッセイは、がんに関与していることが知られている遺伝子 bcl-2 を導入したマウスの組織から、開発され得る。アッセイされ得るヒト乳がん細胞株は、MCF-7、NCI/ADR HS578T、MDA-MB-22231/ATCC、MDA-MB-4335、MDA-N、BT-549、T-47D (NCI、ATCC) である。他の細胞株と組織も使用され得る。生物学的アッセイの限定されない試料を表1に示す。

【0146】

(表1) 細胞株、ヒト組織生検、およびホストに移植されたヒト組織生検 (例えば、ヌードマウス)

病原メカニズム	生物学的アッセイ[乳房、結腸、肺、および前立腺の細胞株 (例えば、乳癌、MCF-7、NCI/ADR、HS578T、MDA-MB-22231/ATCC、MDA-MB-4335、MDA-N、BT-549、T-47D)]
アポトーシス	リガンドと1.5時間のインキュベーション、続いて、FITC Annexin V 染色、すなわち、DAPI染色核形態学的確認。
壊死	リガンドと8時間のインキュベーション (ヌードマウスにおいて)、すなわちヨウ化プロビジウムまたはTOTO-3で生体色素染色、MTTアッセイで確認。
増殖	リガンドと2時間のインキュベーション、続いて、FITC抗PCNAで染色、BRDUで確認。
血管形成	ヌードマウスで腫瘍をリガンドとインキュベーション、フルオレセイン因子V111関連抗原で染色し上皮細胞密度を測定、培養ヒト皮膚微小血管系上皮細胞のβ-tubulinへの移動を確認。
炎症	リガンドと2時間インキュベーションし、TNF, INF, IL-1, IL-2, IL-10, TGFβ, VCAM, ELISAによりNκBを測定。
浸潤	マトリゲル細胞浸潤房でCSFE色素で標識化した細胞を30時間インキュベーション、ヌードマウス試験で確認。
線維症	リガンドと48時間インキュベーション後、フィブロネクチンELISAアッセイまたは免疫組織化学法の実施。
代謝	インスリンとリガンドで2時間インキュベーションし、続いて、グルコース濃度を測定、すなわち、3T3-L1脂肪細胞株およびL6単球細胞株の試験後に、11型糖尿病を正常患者の脂肪生検と比較。
発生/分化	リガンドをMHCクラスII-陰性細胞または単一の多能性ML-IC細胞のいずれかとインキュベーションし、Inaba K et al., 1993, PNAS 90:3038またはPunzel M et al., 1999, Blood 93:3750に従う細胞学的および免疫学的技法による細胞運命の評価。

【0147】

6.1.3.1. アポトーシス

アポトーシスは、細胞膜ホスファチジルセリン結合色素 (FITC Annexin V、Cy5.5 のような別の色素も使用可) によって測定される。結合アッセイで同定された各タンパク質用を選択されたリガンドについて、色々な細胞株に対するアポトーシスが試験され得る。2x10⁵~2x10⁶個の細胞を96ウェルプレートの各ウェルに入れ、1 μM~10 μM 濃度の各リガンドを含む培養液がウェルに3ウェルずつ加えられる。少なくとも、陰性 (リガンドなし) および陽性 (bcl2 反応性リガンド) 対照も実施される。1.5 時間後、FITC Annexin をウェルに加え、15分間細胞と共にインキュベートし、3回洗浄後、蛍光強度がプレートリーダーによって測定される。

【0148】

アッセイは、bcl-2 遺伝子導入マウス (Charles River) のbcl-2 発現細胞と組織を使用して、細胞から組織への移行が可能であることを実証し得る。アポトーシスを誘発するリ

10

20

30

40

50

ガンドを、乳癌患者の採取直後の腫瘍生検によって試験し得る。一次組織生検を使用する利点は、組織採取後2時間以内、例えば組織が虚血に原因する変化を示す前に、アッセイが実施できることである。腫瘍生検の小片が96ウェルプレートに加えられ、上述と同じアッセイが各試料について2ウェルで繰り返される。蛍光強度を読みとった後、試料はDAPI (Molecular Probes, Eugene Oregon) 染色法によって染色され、細胞核の形態、つまり核の縮合および細分の確認のため、蛍光顕微鏡下で評価され得る。また、従来のTUNEL (末端デオキシヌクレオチド転移酵素の仲介によるビオチン化デオキシウリジン3リン酸ニック末端標識) 法がDNAらせん構造の切断部を標識化するのに使用され得る。

【0149】

6.1.3.2. 増殖

細胞増殖は、増殖細胞核抗原 (PCNA) に結合する、フルオレセインで標識化された抗PCNA抗体 (例えば、PC-10、Santa Cruz Biotechnology) に、細胞を曝露させることによりアッセイされ得る。結合アッセイで同定した各タンパク質用を選択されたリガンドについて、細胞株の増殖作用が試験され得る。 $2 \times 10^5 \sim 2 \times 10^8$ 個の細胞が96ウェルプレートの各ウェル中に加えられ、各リガンドを $1 \mu\text{M} \sim 10 \mu\text{M}$ 含んだ培養液がウェルに3ウェルずつ加えられ得る。少なくとも、陰性 (リガンドなし) および陽性対照も実施される。2時間後、FITC抗PCNAをウェルに加え、15分間細胞と共にインキュベートし、3回洗浄後、プレートリーダーを用いて蛍光強度が測定され得る。PCNAアッセイ法は細胞および組織中ですでに使用されている (Kulldorff Mら、2000、J. Clin Epidemiology 53:875)。増殖を阻害するリガンドは、乳がん患者の採取直後の腫瘍生検によって試験され得る。腫瘍生検の小片が96ウェルプレートに加えられ、上述と同じアッセイが各試料について2ウェルずつ繰り返えされる。蛍光値を読みとった後、試料は、蛍光顕微鏡下で評価され、その増殖が本当に影響を受けている細胞が、がん細胞であることを確認し得る。細胞増殖を測定する第二のアプローチは、BRDUまたは ^3H -チミジン摂取を観察する従来法である。第三のアプローチによれば、細胞はCSFE色素 (5, 6 カルボキシフルオレセイン2酢酸サクシニミジルエステル) で標識化され得る。細胞が7 から 8 世代にわたり増殖すると、色素は希釈される。第四のアプローチでは、蛍光を使用したAttoPhosアッセイを利用し、内因性酵素酸性ホスファターゼを測定し、細胞数を測定する。増殖期決定のための7-ADD (7-アミノアクチノマイチンD)、またはKi67抗体による染色を含む、他の方法で増殖中の細胞を検出し得る。

【0150】

6.1.3.3. 壊死

壊死を検出する方法には、ヨウ化プロピジウムまたはTOTO-3のようなDNA結合色素による従来法があるが、これだけに限定されない。またミトコンドリア酵素の遊離を測定するメチルチアゾールテトラゾリウム (MTT) 比色分析を使用して、細胞の生存度を決定し得る。本発明の好ましい実施例では、細胞生存度はDNA結合色素ヨウ化プロピジウムおよびTOTO-3で測定される。

【0151】

細胞株のこれらのアッセイを実行すると、壊死とアポトーシスを区別でき、このアッセイはまた、広く細胞毒性を示すリガンドと特異的な効果を持つリガンドとを区別するのにも役立つ。この区別は壊死とアポトーシスアッセイを同時に実行することで容易になり得る。結合アッセイで同定した各タンパク質用を選択したリガンドについて、細胞株の壊死作用が試験され得る。 $2 \times 10^5 \sim 2 \times 10^8$ 個の細胞を96ウェルプレートの各ウェル中に加え、各リガンドを $1 \mu\text{M} \sim 10 \mu\text{M}$ 含んだ培養液をウェルに3ウェルずつ加える。少なくとも、陰性 (リガンドなし) および陽性対照も実施される。8時間後、ヨウ化プロピジウムまたはTOTO-3をウェルに加え、15分間細胞と共にインキュベートし、3回洗浄後、蛍光プレートリーダーを用いて蛍光強度が測定され得る。

【0152】

壊死は、組織生検に移行させるのが難しいアッセイと思われる。その理由は、一般的に壊死は、少なくとも8時間後に測定され、その時点で組織生検中には虚血により多数の壊死

10

20

30

40

50

が発生し、それが時間を経て高いバックグラウンドとして表れるからである。この問題を解決するために、ヒト生検組織をヌードマウスに移植し、その結果8時間のアッセイ中に起こる虚血により誘発される壊死を防ぎ得る。ヌードマウスにおける成長が腫瘍を変化させないことを確認するために、1ヶ月間ヌードマウスで成長した腫瘍を外植し、上述したように短期間のアポトーシスおよび増殖を試験し得る。腫瘍を組織学的に観察し、採取直後の腫瘍外植部と比較し、その差を評価し得る。同じ標的分子と結合し、50%の例で壊死を起こすリガンドを動物の腫瘍に注入し、8時間後に採取し、ヨウ化プロビジウムで染色し得る。組織学的検査は、腫瘍細胞では壊死が進行中であり、他の生検細胞では壊死は起きていないことを示し得る。

【0153】

6.1.3.4. 血管形成

血管形成の促進または抑制効果を試験するためにインビトロアッセイを使用し、培養したヒト皮膚微小血管内皮細胞の、 β -FGFまたはウシ血清アルブミン（陰性対照）への移動を、阻害対照のアンジオスタチン濃度を増加し、また別なウェル中のリガンド濃度を増加することにより、測定する（Clonetics, San Diego; Polverini PJら、1991、Methods in Enzymology 198: 440）。血管形成もまた長期にわたるイベントであり、そのためヒト生検モデルではヌードマウスでの成長が絶対に必要である。将来、抗血管形成作用をもつリガンドが発見されるなら、腫瘍中へ毎日3-5日続けて注入し、その後その腫瘍を取り出し、抗原に関連する蛍光抗因子VIIIで染色し、内皮細胞密度を測定してそのリガンドをアッセイできるようになる。

【0154】

本発明では、血管形成の他のモデルも意図している。インビボモデルでは、それ上に試験分子を持つヒドロペレットを無血管ラットの角膜に移植する（角膜マイクロポケットアッセイ）。7日間の血管縁からペレットへの増殖を、タンパク質Aビーズ上の抗体により血管形成または抗血管形成タンパク質を取り除くことにより無効になる陽性反応としてそのスコアを付ける（Poverini PJら、1991、Methods in Enzymology 198: 440）。これらの血管の密度、長さ、および管腔サイズの特徴を測定できる。同様なアッセイが、マウスの眼においても実施できる（L Smith, Children's Hospital, Boston）。後肢虚血させたウサギのモデルを使って、血管形成分子を、インビボで試験できる（Shyu KGら、1998 Circulation 98:2081）。他のインビトロ組織モデル系では、未成熟な毛細血管に似ている管構造を形成する3次元培養の内皮細胞を含んでいる（Springhornら、1995、In vitro Cell Dev Biol Anim 31, 473; Sierra-Honigsmann MRら、1998、Science 281:1683）。平滑筋細胞補充は、抗平滑筋アクチンの免疫組織化学によって測定できる。

【0155】

6.1.3.5. 浸潤

腫瘍の浸潤は、マトリゲル（Matrigel）細胞外基質で被膜されたチャンバーである、細胞基底膜浸潤チャンバーを使用してアッセイされ得る。細胞外基質が使用するウェルを被覆し、24ウェルプレート中で1つのチャンバーと他方のチャンバーから分ける（Becton Dickinson Labware）。結合アッセイで同定した各タンパク質用を選択されたリガンドについて、細胞株の浸潤作用が試験され得る。CSFE色素で標識化された細胞はFACSで測定されるか、インビボで細胞運命を追従するのに使用される。また、細胞は、 ^3H -チミジンまたは他のマーカーで標識化できる。約 2×10^5 個の標識細胞を各ウェルに加え、 $1 \mu\text{M}$ または $10 \mu\text{M}$ の各リガンドを含む培養液を、ウェルの上から半分3ウェルずつ加える。 CO_2 培養器でインキュベーションを30時間行った後、膜チャンバーの両側をDMEM/0.1%BSAで3回すすぎ、上部表面を綿棒でこする。ウェル底部にある色素量は、蛍光プレートリーダーを用いて定量され得る。陽性ウェル中の膜を切り取り、底部の細胞数を計測できる。このインビトロアッセイ中で腫瘍浸潤に影響を与えるリガンドは、ヌードマウスのヒト腫瘍生検の組織学分析によって、さらにインビボで試験され得る。

【0156】

6.1.3.6. 発生および/または分化

細胞、組織、臓器または生体の発生および/または分化に対するリガンドの作用を試験するために、様々なアッセイが思索されている。限定のない実施例では、主要組織適合遺伝子複合体 (MHC) クラスII陰性細胞か、単一多分化能骨髄リンパ開始細胞 (ML-IC) のいずれかと共にリガンドをインキュベートし、さらに Inaba K ら、1993、PNAS 90:3038または、Punzel M ら、1999、Blood 93:3750 に従った細胞学的および免疫学的方法により細胞発生運命が評価される。

【0157】

6.2. 実施例2: 糖尿病

末梢インスリン抵抗性は、II型糖尿病を引き起こす主な発病メカニズムであり、疾病による死亡原因の第4位であり、盲目、腎不全および切断の主要原因でもある。インスリンは、筋肉および脂肪細胞におけるグルコースの取り込み、肝および筋肉細胞におけるグリコーゲンの合成、脂肪細胞および肝細胞による脂肪生成、および肝細胞におけるグルコース生成の抑制などを促進させる。NIDDM (インスリン非依存性糖尿病) は、骨格筋および脂肪細胞へのインスリン刺激グルコース取り込みの障害、肝における糖新生の阻害障害およびインスリン分泌調節障害の可能性などの特徴を有する。その経路は、部分的にしか分かっておらず、末梢インスリン抵抗性の原因となる分子は知られておらず、こういった状況は本発明の方法を適用するの適している

【0158】

インスリンは、そのインスリン二量体レセプタの α サブユニットに結合して、レセプタの細胞質ゾルの β サブユニットチロシンキナーゼ活性を誘発するように自身分子および近辺のタンパク質をリン酸化する。インスリンは、DNAおよびタンパク質合成、同化代謝経路の活性化および異化代謝経路の阻害を起こす。IRS-1, IRS-2, IRS-3, IRS-4, Gab-1 および p62 dok など一連のタンパク質すべてが、リン酸化インスリンレセプタと結合し、その基質となり得る。IRS-1 は、一番多くレセプタに関与しているように見えるが、これらのすべてが、ホスファチジルイノシトール3キナーゼの活性化物質であり、横紋筋/脂肪組織特異的グルコーストランスポーター GLUT 4 を、細胞質のゴルジ体から原形質膜に輸送させる。原形質膜では、グルコースが輸送され、その後ヘキソキナーゼによりリン酸化される。(Glut 2 は肝臓と膵臓の β 細胞に存在する)。インスリンはまた、グルコースをグリコーゲンに変換する最終段階を触媒するグリコーゲン合成酵素を上方制御するが、このシグナル経路の前半で欠損が起こると考えられている。

【0159】

肝臓および筋肉がグルコース代謝のほとんどを占めるので、本試験ではこれらの臓器からの細胞を使用する。糖尿病患者の筋肉生検は、健常人の筋肉生検と同様に、インスリンおよび/またはグリクラジドで刺激され得る。ここで健常人はおそらく患者の親族で、そのうち数人は糖尿病の顕在性症状を示さず、インスリンには完全に正常反応を示している。インスリン作用の欠損では顕在性疾患が先行し、糖尿病患者の非糖尿病親族に見られる。示差ディスプレイ cDNA ライブラリが、糖尿病患者および健常人から調製され得る。第二の示差ディスプレイ cDNA ライブラリが、インスリンおよび/またはグリクラジドで刺激された患者生検と、健常人の生検から調製され得る。これらの cDNA ライブラリは、続いてタンパク質として発現され得る。発現したタンパク質と結合するリガンドは、本発明で記載されている方法によって、単離され得る (例えば、HPLC/質量分析)。

【0160】

リガンドは、インスリン刺激後のグルコース取り込み作用について、アッセイされ得る。3T3-L1 脂肪細胞および L6 筋細胞株 (ATCC) は、グルコース代謝の細胞モデルとして使用され得る。 $2 \times 10^8 \sim 1 \times 10^{10}$ 個の細胞を 96 ウェルプレートの各ウェル中に加え、既知濃度のグルコースと各リガンドを $1 \mu\text{M} \sim 10 \mu\text{M}$ 含んだ培養液を 3 ウェルずつ加え得る。少なくとも、陰性 (インスリンなし、リガンドなし) および陽性 (インスリン、リガンドなし) 対照も測定される。次に低濃度および高濃度のインスリンがウェルに加えられる。 CO_2 培養器でインキュベーションを 2 時間した後、グルコース値がグルコース計で測定され得る。細胞株内でのインスリン刺激後、グルコース代謝に影響を与えたリガンドが、II型糖尿

病患者の新鮮な骨格筋および脂肪組織生検を用いて、同じアッセイから試験され得る。組織生検より懸濁化した細胞が96ウェルプレートのウェル中に同じ密度で加えられ、各試料について上述と同じ方法が2ウェルずつ繰り返される。仮にリガンドが末梢インスリン抵抗性を減少させた場合、そのリガンドと遺伝子の組み合わせが、末梢インスリン抵抗の治療法として有用な標的となり得る。その標的は今後さらに試験され、インスリンの代謝的シグナル経路においてマップされ得る。

【0161】

6.3. 既知遺伝子の分子経路内標的の同定

上述のアプローチを使用して、多能性分泌タンパク質のシグナル経路内にある未知遺伝子の機能を同定確認し、組織特異的な方法で毒性作用から治療的应用効果を取り出し得る。TGF β 1は、多数の細胞種における強力な成長阻害物質としてよく知られており、またタイプII TGF β レセプタ、Smad2またはSmad4は、多数の癌細胞で変異することが知られている(Kim SJ, 2000, Cytokine Growth Factor Rev. 11: 159)。一部の腫瘍抑制遺伝子(DPC4)はSMADファミリーのメンバーであり、T細胞免疫応答の強力なダウンレギュレーターである(Prud'homme GJ, 2000, J. Autoimmun. 14:23)。この成長阻害およびアポトーシス誘発経路調節を使用して、がん細胞の成長を阻害し、自己免疫中にT細胞の耐性を誘発し、TGF β 経路の遮断によってがん抗原への耐性を破壊するような新規な治療薬が開発され得る。

【0162】

こういった開発の限界要因の1つは、TGF β 1が細胞外基質の析出を誘発することであった(Massague, 1990, J Ann Rev Biochem 6:597.)。その析出には、フィブロネクチン、コラーゲン、プラスミノゲン活性化因子インヒビター-1および基質メタロプロテアーゼの組織インヒビターを上方制御し、間質のコラゲナーゼのような基質分解プロテアーゼを下方制御する過程が含まれる。基質成分の過剰生産は、組織線維症における主な所見であり、腎臓および他の疾患を末期症状に導く重要な原因である。(Blobe GC, 2000, NEJM 342: 1350)。フィブロネクチン生成の抑制はがんによく観察され、細胞接着を減少させ、転移を増大させる原因となる(Kornblihttら、1996、FASEB J 10:248)。c-jun N-末端キナーゼ(MAPキナーゼのファミリーメンバー、JNK)がcJUN(転写因子のAP-1ファミリーメンバー)およびATF-2(別の転写因子)を調節するように活性化されるSmad非依存性経路を通して、TGF β は、これらの作用をECM(細胞外基質)上で誘発させる(Hocevarら、1999、EMBO J 18:1345)。TGF β の多能性効果はjunおよびsmad経路を別々に標的とすることによって、詳細に分析され得る。このために、ヒト一次T細胞および線維芽細胞を2分し、それらの細胞の半分が、アンチセンスjunまたはSMADを含むレトロウイルスで形質移入され得る。また、この移入は、異なるベクターで達成され、細胞がsmadまたはjunに反応性のあるペプチドで形質導入され得る。その結果得られた細胞株は、続いてTGF β によって刺激され、刺激を受けた細胞と受けない細胞間で示差的に発現されと思われるcDNAがクローン化され、いずれかの経路を持つ細胞が、マイクロアレイ解析または他の示差発現技法によって遮断され得る。一旦cDNAが1つの経路のみのに関連する発現として同定されると(しかし、その機能は未知)、これらのcDNAは、タンパク質として発現され、そのタンパク質に結合するリガンドは、生化学的結合アッセイおよびHPLC分析と質量分析によって、単離され得る。細胞外基質の急増(上述のPCNAに基づいたアッセイで)または分泌を遮断あるいは誘発する能力について、引き続きリガンドを試験できる。細胞外基質アッセイにおいて、48時間内で細胞外基質の主な成分であるフィブロネクチン沈着物が、96ウェルプレート上でフィブロネクチンELISAアッセイを用いて測定されるであろう。この方法で、遺伝子を同定でき、タンパク質の抗急増作用に関連するが、線維化促進作用には関連しない標的、またその逆の標的を確認することができる。同様なアプローチを使用して、細胞または組織刺激物質を観察し、分子経路の新たなメンバーを同定し、それらを薬物標的として確認し得る。

【0163】

7.1. 表現型から遺伝子型へ

10

20

30

40

50

7.1.1. 表現型の検出

6.1.3.1 項および6.1.3.2. 項 で記載されている腫瘍細胞のアポトーシスおよび増殖アッセイを、例えば384ウェルプレート型を用いてハイスループットスクリーニングに適用し得る (Applied Biosystems FMAT 8100)。アポトーシスおよび壊死を同時にアッセイし得る。アポトーシスおよび壊死については、Cy5.5 Annexin V アッセイおよびTOTO 3 試薬をそれぞれ使用し得る (Applied Biosystems)。Cy5.5 標識化抗PCNA抗体 (PC-10、Santa Cruz Biotechnology) を、細胞増殖アッセイに使用し得る。アッセイされ得るヒト乳がん細胞株の限定されない例として、MCF-7、NCI/ADR HS578T、MDA-MB-22231/ATCC、MDA-MB-4335、MDA-N、BT-549、T-47D (NCI、ATCC) が挙げられる。アッセイされ得るヒト前立腺がん細胞株の限定されない実施例は、DU-145、PC-3、LNCaPである。アッセイされ得るヒト大腸がん細胞株の限定されない例として、COLO 205、HCC-2998、HCT-15、HCT-116、HT29、KM12、SW-620が挙げられる。アッセイされ得るヒト肺がん細胞株の限定されない例として、A549/ATCC、EKVX、HOP-62、HOP-92、NCI-H23、NCI-H226、NCI-H322M、NCI-H460、NCI-H522が挙げられる。1x10⁵~1x10⁸個の細胞を384 ウェルプレートの各ウェル中に加え得る。1 pM~1 M、好ましくは1 μM~10 μM のリガンドライブラリ中の有望な各リガンドを含む培養液 (上述5.1.2 項でリストされている限定されない例) を試験するウェル中に3ウェルずつ加える。陰性 (リガンドなし) および陽性 (スタウロスポリン) の対照を含む。≤20 μM で表現型の作用を有するリガンドは、本発明によれば標的同定の優れた候補分子である。

【0164】

7.1.2. 標的物質の同定

本発明の重要な利点は、従来技術と異なり、1つ以上の生物学的アッセイで作用を持つことが認められたリガンドの標的が、そのリガンドを用いて同定し得ることである。本発明によれば、標的を同定するのに使用し得るアプローチは多数存在する。

【0165】

最初の一連の態様では、可能性のある標的は、細胞表面上に表現されるタンパク質である。1つの限定されない例によれば、全長ヒトcDNAのライブラリは、pDisplayベクター (Invitrogen) 内で発現される。このベクターはタンパク質を標的とし、それを真核細胞表面上の細胞膜に固着させる。本発明の別の限定されない実施例では、全長ヒトcDNAライブラリは、pYD1酵母ディスプレイベクターまたはEBY100 出芽型酵母菌 (*Saccharomyces cerevisiae*) 株に形質移入された類似ベクター内で発現される。さらに本発明の別な限定されない実施例では、全長ヒトcDNAライブラリは、バキュロウイルスベクターによる昆虫細胞の表面上で発現される (Ernst W ら、1998, *Nucleic Acids Research* 26:1718)。ペプチドの発現のみ起こさせる原核生物系とは反対に、これらの系では、全長タンパク質を表面上で発現させる段階ができる。

【0166】

代替的態様では、ポリヌクレオチドライブラリを、ペプチド単独で、または1つの細胞またはウイルスの表面上の融合 (例えば、バクテリオファージT7またはM13) として発現できる。限定されない例は、ヒトまたは感染物質から作製されたポリヌクレオチドライブラリを含む。本発明の特異的な実施例で、cDNAは、pFliTrxベクター (Invitrogen) または類似ベクターにおいてドデカペプチドとして発現される。この実施例によれば、ベクターが大腸菌中で発現されると、ペプチドはチオレドキシタンパク質の活性部位ループおよび細菌フラジリン遺伝子内でディスプレイされる。本発明の別の実施例では、プロマイシンの処理により、ペプチドがそれをコードしているRNAに融合されるリボソームディスプレイ系で、有望な標的をペプチドとしてディスプレイし得る (Robert R W ら、1977, *PNAS* 94:12297)。本発明に従って、他のすべてのディスプレイ系 (レトロウイルス、アデノウイルスなど、しかしこれだけに限定されない) を利用して、cDNAまたはペプチドをディスプレイし得る。

【0167】

7.1.3. 分離

10

20

30

40

50

上述のいずれの方法によってディスプレイされた有望な標的物質をリガンドに曝露し得る。リガンドを、表面、ビーズまたはカラム上に固定化されるか、または使用する分離方法に依存して溶液中に存在し得る。本発明の最初の実施例で、リガンドを、表面上で直接固定化、直接標識化、または検出され得る。本発明の第二の実施例では、これ以前の例で図解されているように、標的分子がディスプレイされているリガンド-標的分子対の採取を容易にするように、リガンドを親和性標識物質で誘導体化させ得る。このような親和性標識物質の限定されない例として、ビオチン、ジゴキシゲニン、または抗体が含まれる。リガンドと結合するディスプレイされた標的分子を、結合していない標的から分離し、その標的をコードする配列を標準クローン化およびDNA配列決定法によって同定し得る。

【0168】

本発明の最初の態様で、細胞は、蛍光標識化されるかまたはビオチン化されたリガンドで「染色」され（後者は、FITC アビジンと結合）、フローサイトメーター（MoFlo HTS Cytometer, Becton Dickinson FACS）でプレートのウェルまたはチューブなどに分別される。この細胞は続いて、標準の細胞培養方法で増殖され得る。最初の限定されない例によれば、薬物レセプタをコードする遺伝子は、COS 1 細胞のプラスミドの回収から、複製のSV40 起点に対する大型T抗原作用の効果を利用して、クローン化され得る。第二の限定されない例によれば、PCRがプラスミド挿入物を回収するために使用され得る。

【0169】

本発明の第二の態様では、細胞、ウイルス粒子またはペプチド-ヌクレオチド融合物が、薬物被覆磁気ビーズ、薬物被覆表面（例えば、パニング用ウェル）または薬物被覆カラムによって選択され得る。表面、ビーズ、またはカラム上の薬物リガンドは、低親和性相互作用における結合活性を増加させるために、高密度であることが好ましい。薬物は、親和性標識化合物（例えば、アビジン、ジゴキシゲニン）と通じて表面、ビーズ、またはカラムに付着し、1回以上の洗浄操作後溶出を達成し得る。磁気ビーズの場合、マグネットを使用し、洗浄中にビーズを単離し、結合細胞、ウイルス粒子またはペプチド-ヌクレオチド融合物を回収し得る。パニングの場合、ウェルで保持される細胞、ウイルス粒子またはペプチド-ヌクレオチド融合物に対する各連続洗浄操作を完了後、上清を注ぎながら捨てる。カラムの溶出は、標準的な方法で達成し得る。リガンドを親和性標識物質で誘導体化した場合、細胞、ウイルス粒子またはペプチド-ヌクレオチド融合物は、遊離の親和性標識物質をカラムに過剰に加えることにより、カラムから溶出し得る。

【0170】

標的発現細胞またはウイルスは一旦分離すると、適当に増殖できる。続いて、標的分子をコードするcDNAを標準的な分子生物学方法によって回収し得る（例えば、プラスミド回収またはPCR）。精製されたペプチド-ヌクレオチドの場合、cDNAの部分的配列がRT PCRによって、同定されるであろう。上述のアプローチによって、1回以上の選択を行い、標的を精製し、クローン化できる。この方法で、今まで未知であった薬物標的をコードするDNA配列を単離でき、その薬物標的をコードするcDNAをクローン化するのに使用できる。

【0171】

薬物標的をコードするcDNAを一旦同定すると、6.1項のように、疾病組織の細胞内の示差的発現を試験するのにcDNAを使用できる。標的が疾病と正常細胞間で示差的に発現されると、特異性が確立し、標的と相互作用するリガンドを、その疾病に対するインビトロおよびインビボの生物学的アッセイで試験し得る。

【0172】

そこで、表現型アッセイで機能に関与する標的を本発明を用いて同定する。

【0173】

7.2. プロテオミクスによる標的分子の同定

また、対象リガンドを複数の有望な標的と組み合わせるために6.1.2項で述べた方法を用い、リガンド-標的対を採取し、リガンドおよび標的を随意に解離させることによって、標的の同定を達成し得る。続いて、標的を同定し得る。本発明の1つの実施例で、標的は、一般的な方法（例えば、アミノ酸配列決定、質量分析および/またはNMR）によって同定

10

20

30

40

50

され得るタンパク質である。タンパク質を一旦同定すると、標準のプロテオミクスによって疾病細胞との関連性を決定し得る。

【0174】

8.1. シグナル経路のマッピング

一旦、幾つかの遺伝子が発病の特定な分子経路に関与することが示されると、その標的成分を、他の分子経路の構成成分に関連して分子経路内でマッピングできる。異なる分子経路の構成成分に結合するリガンドは、光活性化架橋剤によって誘導体化され得る。少なくとも1つの既知分子経路の構成成分が、GFPのようなマーカーで融合される。以下示す物質をインビボおよびインビトロで組み合わせ得る。(i) 既知分子経路の構成成分と結合する誘導体化リガンド、(ii) マークした経路の構成成分、例えば、GFP融合タンパク質、(iii) 別の分子経路の構成成分と結合あるいは結合する可能性のある、少なくとも1個の誘導体化リガンド、および(iv) 他の分子経路の構成成分。架橋を起こす刺激物質を適用して、得られる複合体の各成分を同定する。この方法で、各分子経路の構成成分を、分子が相互作用する他の成分に関連してマッピングし得る。本発明のさらなる利点は、経路エフェクターがこの方法によって同定され得ることである。さらに、各経路の構成成分のプロファイルが、もしあれば、その経路を通じて作用する既知薬物と比較し、さらに比較試験をその発病経路によって起こされる異なった疾病の細胞に基づいたアッセイで行うことが出来る。この情報は、特定の疾病適応に対しベストな標的を確認し選択するために使用できる。また代わりに、この情報は、特定な患者に対し、薬理遺伝学によるベストな治療法を選択する場合にも使用し得る。

【0175】

9.1. リード化合物の最適化

多数のリード化学物質が生化学的および表現型レベルで特徴付けられるので、構造活性相関(SAR)が確証され、リード化合物の最適化の基礎となり得る。類似活性を持つ2-3個の分子が同定されると、アッセイでその構造と活性を比較することによりSARを決定できる。標的に的を絞った合成技術を使用して、お互いに近くで結合している分子を架橋することができ、その結果、結合分子の作用が、タンパク質の同じサブサイトまたは、標的タンパク質の異なるサブサイトを通して仲介されるかを示すことが出来る。この方法で、標的上で追加される別の機能を持つサブサイトがマップされ、異なるメカニズムを、これらのサブサイトで結合する分子による表現型の結果から解釈することができる(例えばアゴニスト対アンタゴニスト)。

【0176】

標的分子に的を絞った合成の第二の利用法は、その標的に対するリガンドの親和性を増加し、その結果リガンドを、表現型を遺伝子型にリンクさせるのに役立て、より有効な薬物リード化合物にすることである。リガンド構造骨格上の1つの官能基にある光活性化架橋剤を使用して、標的と結合しているリガンドをリンクさせ、その結果標的分子をテンプレート(鋳型)として使用し得る。このリンクにより、標的分子への結合親和性が少なくとも2~10倍増大し、さらに標的分子に的を当てた生体関連方法でライブラリの構造多様性を強化することになる。

【0177】

10. 表現型と遺伝子型をリンクするシリカ(IN SILICA)によるアプローチ

当該発明は、表現型を遺伝子型に関連づけるようにシリカ(in silica)でマッチする各リガンドまたはリガンドセットに対する、リガンド-標的(遺伝子型)およびリガンド-生物学的アッセイ(表現型)の化学的フィンガープリントを確立する方法を提供する。

【0178】

本発明は、リガンド-標的対の実験データを保存する第一情報検索システムを提供する。本発明は、試験した各生物学的アッセイ中の各リガンドの作用を保存する、第二の情報検索システムを提供する。本発明は、各標的の機能および/または発現パターンが、既知な場合、保存される第三の情報検索システムを提供する。これらのシステムは利用し易いように随意に統合され得る。

【0179】

本発明の1つの態様で、システムに入力されたデータは、すべての標的についてリガンドとの結合を試験するか、またはすべてのリガンドについて各生物学的アッセイで試験する、ショットガンアプローチによって得られ得る。例えば、標的セットは、最高全遺伝子までのすべての発現生成物まで、さらに選択した生物体のゲノム中の全遺伝子まで包括し得る。各標的は続いて、リガンドライブラリのスクリーニングを行うのに使用され、結合しているリガンドを同定する。このデータを第一情報検索システムに入力する。

【0180】

別の例によれば、リガンドの大規模組み合わせ化学ライブラリの各メンバーの作用が利用できる各々の生物学的アッセイで測定される。このデータを第二情報検索システムに入力する。

10

【0181】

本発明の別の態様では、特定疾患に対し選択された標的に結合する、または選択した生物学的アッセイで選択されたリガンドによって誘発された表現型に結合する、リガンドに的を絞った解析によって、システムに入力されたデータが得られる。このデータを必要に応じて第一または第二情報検索システムに入力する。

続いて、これらのシステムを使用して、示差的発現データがないか、または特定の疾患に焦点を当てない場合でも標的機能を予測できるように導き得る。さらに、これらのシステムはユーザーを、特異的な作用を持つリガンドおよび標的を選択するように導き得る。さらなる利点は、本システムは必要な結合に関する実験および生物学的アッセイの数を減らし得ることである。他の利点は、熟練した技術者には明らかである。

20

【0182】

本発明の1つの態様で、ユーザーは対象標的を選択する。次に、ユーザーは、対象標的に結合するリガンドを、実験的にまたは第一情報検索システムから同定する。続いてユーザーは、同定したリガンドについて第二情報検索システムに検索要求をし、各リガンドに関連する表現型を決定する。この方法で、標的は1つ以上の表現型と関連づけられ得る。

【0183】

本発明の別の態様で、ユーザーは対象表現型を選択する。次に、ユーザーは、選択された表現型を調節するリガンドを、実験的にまたは第二情報検索システムから同定する。ユーザーは、同定したリガンドについて第一情報検索システムに検索要求し、リガンドが結合している標的を同定する。この方法で、表現型を1つ以上の標的と関連づけられ得る。

30

【0184】

本発明の別の態様では、これらの情報検索システムを、標的機能情報および/または発現解析データと組み合わせ、ユーザーが標的と薬物リード化合物を確認できるように導き得る。本実施例の最初の例で、ユーザーはタンパク質である標的XおよびYを選択し得る。Xをコードする遺伝子が正常細胞で発現されるが、腫瘍細胞で発現されないことを示す発現データを、ユーザーは得る。さらに、ユーザーは、Yをコードする遺伝子が正常細胞で発現されず、腫瘍細胞で発現されることを示す発現データを得る。続いて、ユーザーは第一情報検索システムに検索要求を行う。この検索要求の結果を表2に示す。

【0185】

(表2)

40

標的	結合リガンド
X	1
X	2
X	3
Y	2
Y	3
Y	4

10

20

【0186】

続いて、ユーザーは第二情報検索システムに検索要求を行う。この検索要求の結果を表3に示す。

【0187】

(表3)

リガンド	表現型
1, 2, 3	血管形成
2, 3, 4	増殖

30

【0188】

この例によれば、ユーザーは、がんの治療法用の有効な標的物質として標的Yを選択し、特異的にYと結合しXとは結合しない能力を有するリガンド4を選択し得る。従って、本発明は、標的を確認し、薬物リード化合物を同定するようにユーザーを導くことができる。

40

【0189】

本発明の第二の例では、表現型から遺伝子型へのアプローチを使用し、リガンド1、2、および3が生物学的アッセイでアポトーシスを誘発させること、リガンド3、4および5は、血管形成を刺激すること、およびリガンド1、3および6は壊死を誘発することを確認した。この情報は、情報検索システムに保存される。ハイスループット結合アッセイにおいて、リガンド3と4が $K_d < 50 \mu M$ で標的Xに結合することが認められた。情報検索システムによる検索は、熟練技術者に、(i) 標的Xは血管形成に関与し得ること、(ii) リガンド3は薬物リード化合物としては不十分な候補であること、および(iii) リガンド4は薬物リード化合物として優れた候補であることを示すことになる。

【0190】

50

11. 本発明方法の自動化

図18および19に図解したような高度に自動化されたアプローチは、本発明の別な実施例である。このアプローチには、化合物ライブラリからリガンドを決定するには量が不十分な場合でも、発現ベクターのハイスループット構築、タンパク質生成、および1週間に>20のタンパク質の生成を可能にする精製装置が含まれる。これに引き続き、化学アレイアッセイのようなハイスループットアッセイを行い、標的骨格構造の対を同定する。これらの標的骨格構造対は、図17に概略したような利用法を持つ化学アレイデータベースを構成する。

【0191】

発現ベクターのハイスループット構築では、例えば、NCBI, Stratagene, または Incyte より得られるヒトプロテオーム内の1個のタンパク質をコードするcDNAを、96ウェル型の自動化液体処理システム (Tecan) を使用して、DES発現ベクター (Invitrogen) に挿入する。DES発現ベクターは、コードされたタンパク質に分泌シグナルとHisタグを加えるので、その結果、そのコードされたタンパク質は培養液中に分泌され、Hisタグと結合するニッケルカラムを使用して生成できる。ベクターは、続き形質転換受容性な大腸菌内に形質移入され、細胞が増殖される。この発現ベクターを、ロボット操作液体ハンドラーを用いて大腸菌細胞から抽出し、標準の溶解試薬を加え細胞を溶解させ、さらに溶解物をQiagenカラムに適用し発現ベクターを精製する段階ができる。特定な実施例において、QIAwell 96 Ultra Plasmid Kit を使用して溶解物を精製する。このキットでは、溶解物の除去にQiafilter 96 ウェルプレート、プラスミドDNAの精製にはQIAwell 96ウェルプレートを、さらに QIAvac 96 自動吸引装置で各プレートを順番に脱塩するにはQIAprep 96 ウェルプレートを使用している。必要に応じて、適切な読み取りフレーム内でcDNAが挿入されている発現ベクターを含む細胞が、標準的な方法で選択される。例えば、発現ベクターは制限酵素で消化され、あるいはフレーム内にcDNA挿入を含むかを確認するために配列決定される。

【0192】

挿入部を含む発現ベクターは続いて、リン酸カルシウムによる標準的な形質移入方法によってショウジョウバエS2細胞 (Invitrogen) に形質移入され、Select自動化組織培養システム (Automation Partnership) 中で、1ベクターにつき6~12個の フラスコ中のショウジョウバエ発現培養液 (Invitrogen) 内で増殖される。各Selectシステムで、最大150個の フラスコ、すなわち最大40個の異なるタンパク質を発現する細胞株を取り扱うことができ、同時に複数のSelectシステムを使用すると、1週間に600個のタンパク質まで処理量を増大できる。24時間後、硫酸銅を培養液に加え、タンパク質発現を誘発し、3日目と7日目に上清を採取し、Biorobot (Qiagen) 上の96 ウェル型のニッケルカラム (Qiagen QIAexpress タンパク質精製システム) 内に通す。続いて、SDS分析または他の品質管理分析 (Qc) を行うために、Tecan 液体ハンドラーによってこのタンパク質量がPHAST ゲル (Pharmacia) に移される。

【0193】

残りの試料が、試薬保管検索システム (Haystack) によって、化学アレイアッセイ (例えば、本書に記載したいずれかのアッセイ法) および保存用フリーザーに移動する。例えば、ロボット操作液体ハンドラー (Tecan) を使用して、精製された標的タンパク質を候補リガンドのライブラリと組み合わせ、96 ウェルプレートのウェル内で1つ以上の候補リガンドを標的タンパク質に結合させることができる。続いて、標的タンパク質と候補リガンドを含むアッセイ混合物を96 ウェルプレートから注入し、リガンドが結合した標的タンパク質を単離するために同時に最大6 カラムまで運転可能なHPLC (Waters 2790) に、この96 ウェルプレートを移動できる。リガンドが結合した標的を含む画分がフラクションコレクタ (Gilson) によって採取される。代わりとなる実施例では、ロボット操作液体ハンドラー (Tecan) を使用して、精製された標的タンパク質を候補リガンドのライブラリと組み合わせ、96 ウェルプレートのウェル内で1つ以上の候補リガンドを標的タンパク質に結合させる。この96ウェルプレートは、例えば、標的タンパク質を未結合リガンドから

分離させる樹脂を持つカートリッジを含んでおり、ロボット (TecanまたはQiagen) により全量を移動させ、結合リガンドと標的タンパク質を2番目の96ウェルプレート内に単離する。代わりの実施例では、結合が96ウェルプレート内で起こり、続いて液体ハンドラー (Tecan) が試料を、分離用のカートリッジを含む2番目の96ウェルプレートへ移動させる。さらに別な実施例では、カートリッジは多ウェル型 (Pharmacia) で利用できるスピニカラムでもある。チップおよびキャピラリーLCを使用した分離方法も使用できる。洗剤または他の変性剤を液体ハンドラーで (Tecan) 加えて、結合リガンドをタンパク質から遊離させることができ、続いて、遊離リガンドを分析用の適切な機器に加える。例えば、リガンドは、自動インジェクタ (Waters) の付いたHPLCの逆相カラムによって、質量分析計に注入され、MALDITOF質量分析用にフィルター上にスポットされるか、またはNMR、IR、FTIR、またはUV分光計で測定される。代わりの実施例では、結合リガンドを伴う標的タンパク質を、液体ハンドラー (Tecan) によって、96ウェル型MALDITOF質量分析計 (Bruker Daltonics) に装填するか、スポットする。別に代わる実施例では、結合リガンドを伴う標的タンパク質が、ロボット操作による吸引により、96ウェル型プレートのフィルタ (例えば、ニトロセルロース) へ全量移動される (Tecan)。別の実施例では、この同じフィルタ上への吸引を、フィルタをカートリッジと真空装置との間に置き、96ウェルカートリッジの吸引と同じ操作で実施する。続いて、MALDITOF質量分析計により、96個の各スポットから標的タンパク質とリガンドが解離し、化合物および/または複合物の質量スペクトルが生成する。本書で述べた情報システムによるデータ処理後、リガンドとその標的の同定結果が化学アレイデータベース (Chemical Array Database) に入力される。これらの方法のいずれも、384、1536 (個) ウェル、チップ使用、または他の型で実施できる。同様に、いずれのデータも、IDBS Activity BaseまたはPrice Waterhouseに基づく検査室情報管理システム (LIMS) または、他のLIMSソフトウェア/システムによって、入力管理できる。

【0194】

HEK293細胞、CHOまたはCOS細胞を含む、他の一過性発現に基づく生産系に、同様な方法を適用できるが、これに限定されない。また、ローラーボトルシステム、攪拌タンクシステム (例えば、Celligen Plus社、New Brunswick) またはキャピラリー細胞培養システム (Amicon) のような、他の自動化または半自動化生産システムも使用できる。別の実施例では、New Brunswick社の1L以上のバイオリアクターのような半自動段階を使用し、上述のようにベクターのpCDNAファミリー (Invitrogen) に基づいて構築した発現作成物が一過的に形質移入されたHEK293細胞 (Life Technologies) のような細胞を増殖させる。一過的に形質移入されたCHO細胞も使用できる。これらの細胞種の形質移入は、Lipofectamine 2000 (Life Technologies) を使用して効率よく実施できる。代わる実施例において、他の形質移入方法が使用される (例えば、電気穿孔法、リン酸カルシウム、リポフェクチン、Lipofectamine Plus (Life Technologies) または他の標準技法)。これらの細胞は、DMEM、または他の血清含有標準培養液や血清不含の標準方法で増殖される。さらに、Invitrogen社のカタログ、他のベクター企業、科学文献に記載されているような様々な細胞株に適切なベクター、または熟練技術者に明らかなベクターなど、別の発現ベクターも使用される。

【0195】

必要に応じて、クローン化選択操作を実行し、その結果、生産システムに基づく安定な生成細胞株が得られる (例えば、CHOまたは大腸菌ベースのシステム)。代表的なクローン化選択操作には、多ウェル型で、例えばゲネチシンのような選択性抗生物質の存在下、細胞を増殖させ、発現ベクターを含んでいそうな細胞を選択する段階、および標準のELISAアッセイまたはタンパク質中のHisタグを検出するための他の標準アッセイによって、各ウェル中の分泌タンパク質の存在を確認することが含まれる。

【0196】

さらに、ハイスループット生産およびスクリーニング技法は、本発明のいずれの方法にも使用できる。例えば、いかなる結合アッセイ (チップ、フィルタ、放射同位体標識化、蛍

10

20

30

40

50

光、表面プラズモン共鳴法など)、生産方法(例えば、CHO、HEK 293、Cosのような哺乳類の細胞、ショウジョウバエのような昆虫の細胞、大腸菌のような細菌、またはピチア(pichia)のような酵母)、生産システム(例えば、バイオリアクター(Brandel社のNew Brunswick システム、フラスコ使用、細胞キューブ、表面結合、懸濁培養、血清含有培養液、または血清不含培養液)、およびいかなる精製法(Hisタグ/ニッケルカラム、GST/グルタチオン、インテイン、または他の親和性カラム)も使用できる。これらの自動化および/またはハイスループットのいずれの方法も、複数ロボットシステム(Automation Partnership社の複数Selectロボットのような)のようなマルチシステム動作で同時に実行できる。例えば、2、4、5、6、8、10、 10^2 、 10^3 、 10^4 、 10^5 、 10^6 個またはそれ以上の標的を同時にアッセイし、標的に結合するリガンドを選択できる。同様に、2、5、10、 10^2 、 10^3 、 10^4 、 10^5 、 10^6 、 10^7 、 10^8 、または 10^9 個以上の関心対象の小分子を同時にアッセイし、小分子と結合する標的分子を選択できる。

10

【0197】

他の実施例

前述の記載から、様々な利用および条件に本発明を適用するために、本書に記した本発明に変更および修飾を行い得ることは明らかである。そのような実施例は、また、特許請求の範囲内にある。

【0198】

様々な刊行物および特許出願が本書で引用されており、その内容は、各独立した刊行物および特許出願が参考文献として組み込まれることを具体的に個々に示す場合と同じ程度に、本明細書に参考としてその全体が組み込まれる。

20

【図面の簡単な説明】

【0199】

【図1】「遺伝子型から表現型へ」のアプローチの概略を示す。

【図2】「表現型から遺伝子型へ」のアプローチの概略を示す。

【図3】マイクロモル単位の親和性を持つ特異的リガンドを単離し抽出する、P38 MAP キナーゼの能力を図解した一連のスペクトルを示す。

【図4】P38 MAP キナーゼ濃度依存性による86002 ピークの減少と、遊離化合物からのタンパク質結合化合物のHPLC分離におけるキニンピークのごく僅かな減少を表す、一連のUVスペクトルを示す。

30

【図5】混合物から抽出され、p38 MAP キナーゼから遊離した化合物が、86002であると同定されたことを表す一連の質量スペクトルを示す。

【図6】10種混合物中の各化合物とその分子量のリストである。

【図7】P38の濃度依存性の86002 ピークの減少と、コルヒチンピークあるいは、遊離化合物からのタンパク質結合化合物のHPLC分離中、混合物中の他の化合物を表すピークの、ごく僅かな減少を明らかにした一連のスペクトルを示す。タンパク質画分を採取し、質量スペクトルを測定すると、そのスペクトルには、他のピークよりもはるかに高い強度に86002 の特徴を示すピークが生じた。

【図8】チューブリン濃度依存性のコルヒチンピークの減少と、86002ピークあるいは、遊離化合物からのタンパク質結合化合物のHPLC分離中、混合物中の他の化合物を表すピークの、ごく僅かな減少を表した一連のスペクトルを示す。タンパク質画分を採取し、質量スペクトルを測定すると、そのスペクトルには、他のピークよりもはるかに高い強度にコルヒチンの特徴を示すピークが現れた。

40

【図9】100種混合物中の各化合物とその分子量のリストである。

【図10】P38 MAPキナーゼが、特異的な濃度依存性様式で、100種の混合物からマイクロモルの親和性を持つ1つのリガンド(86002)と結合し、抽出することを表した一連のスペクトルを示す。

【図11】チューブリンが、特異的な濃度依存性様式で、100種の混合物からのヒット化合物(コルヒチン)と結合し、抽出することを表した一連のスペクトルを示す。

【図12】標的タンパク質を100種混合物中の非結合化合物から、高流速で良好に分離さ

50

れることを表した一連のUVスペクトルを示す。

【図13】非結合化合物から標的タンパク質に結合した化合物を分離する、スピンカラムの能力を表した一連のスペクトルを示す。この方法を使用して、チューブリンに結合した100種混合物から主な化合物としてコルヒチンを同定する。

【図14】化学アレイアッセイの1つの実施例段階を表した図を示す。

【図15】代表的なコンピュータ配置図を表す。

【図16】試料中の化合物同定のために本発明の1つの実施例で使用する代表的なフローチャートを示す。

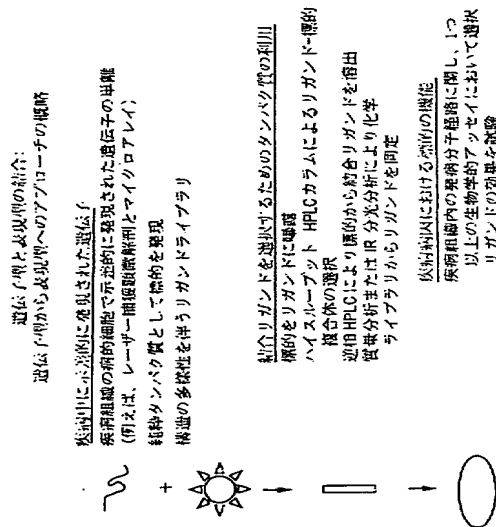
【図17】ヒトプロテオームの化学的フィンガープリントを生成するのに使用され得る、化学的骨格と標的タンパク質との組み合わせを表したグラフを示す。

【図18】リガンド/標的対を生成するための本発明の自動化ハイスループット方法の1つの実施例の図解を示す。

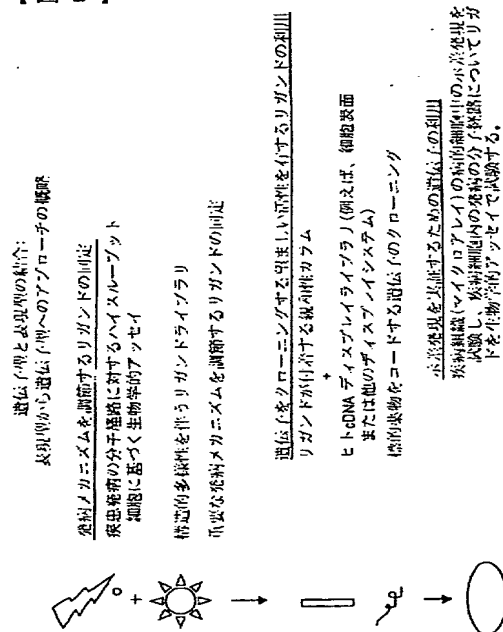
【図19】自動クローニングおよび生成システムによって、1週当たり約600個の割合で約3年間にわたり、ヒトプロテオーム中の約90,000個の各タンパク質を約2 mgハイスループット生産するための1つの実施例の図解を示す。

10

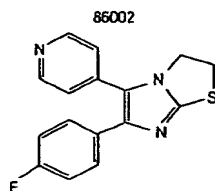
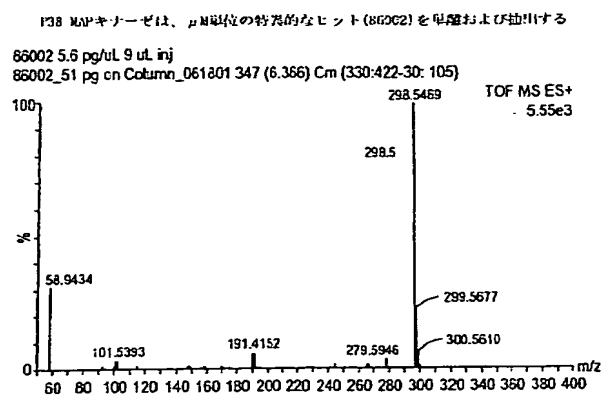
【図1】



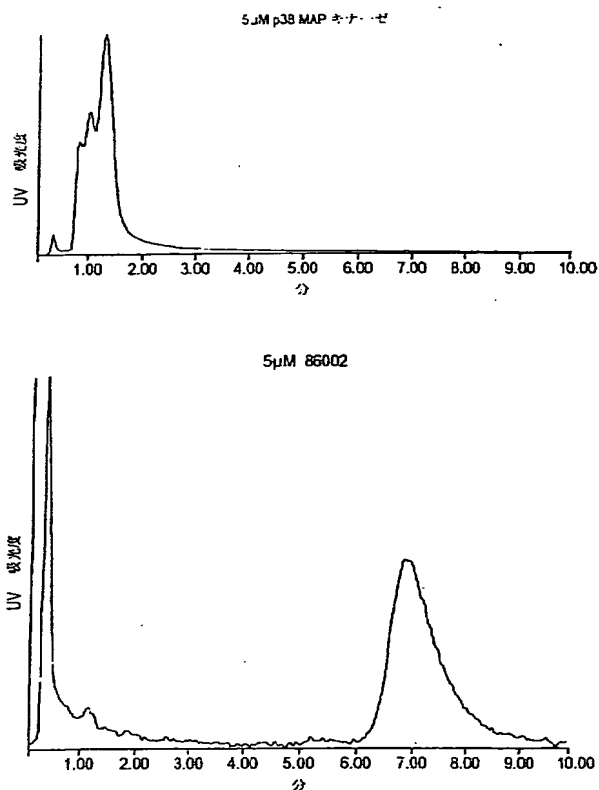
【図2】



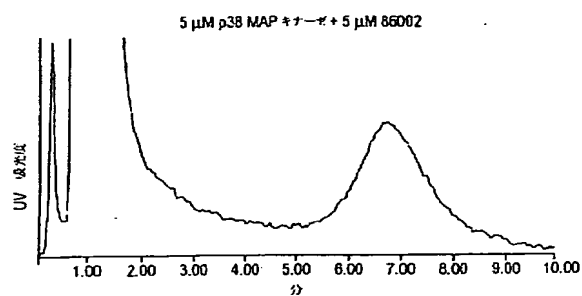
【図 3 - 1】



【図 3 - 2】

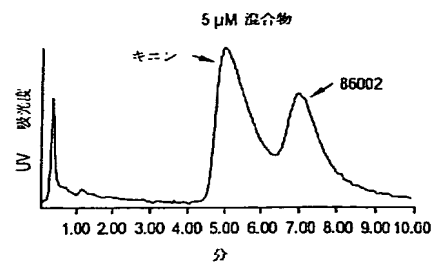


【図 3 - 3】

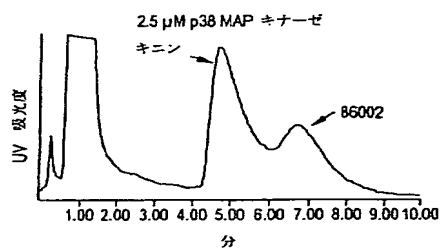
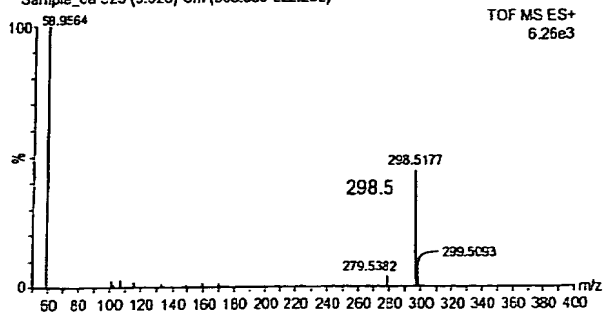


【図 4 - 1】

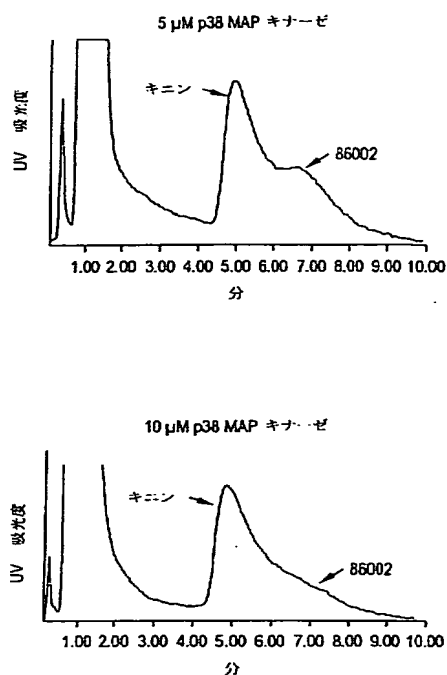
p38 MAPキナーゼは、温度依存性の様式で、非特異的対照 (キニン) ではなく μ M単位の特異的なヒット (86002) に結合し、抽出する



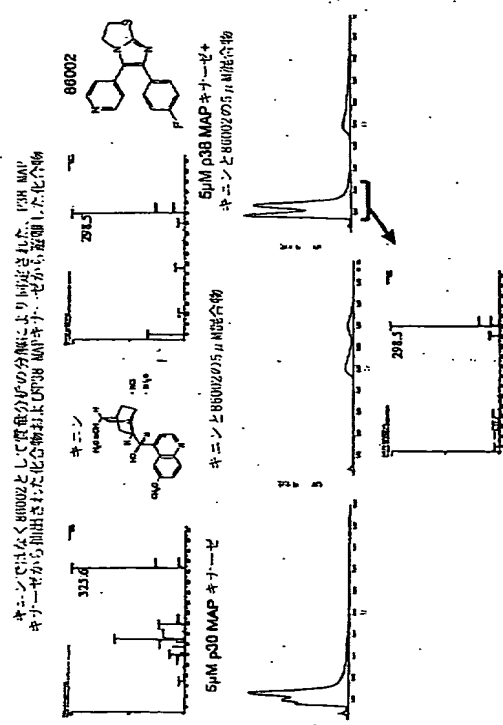
5 μ M p38 5 μ M 86002
Sample_6a 323 (5.928) Cm (305:353-222:232)



【圖 4-2】



【图 5】



【图 6-1】

10種類の化合物の混合物

化合物
ラニチジン

MW
350.9

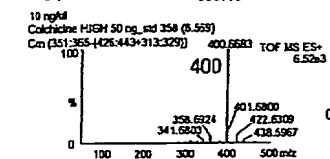
構造式

$$\begin{array}{c} \text{NHMe} \\ | \\ \text{H}-\text{C}=\text{CH}-\text{NO}_2 \end{array}$$

コルヒチン

399.43

HC1

COc1cc2c3c(c1)cc(OC)c(=O)c3c2c4cc(OC)c(=O)c(NC(=O)C)c4

ミノサイクリン

493.9

CN(C)c1c(C)cc2c(c1)c(O)c(O)c(O)c2C(=O)N · HCl · xH₂O

タクリン

234.7

Nc1ccc2c(c1)c3ccccc3n2 ·HCl ·H₂O

ソロソロノロー

295.81

CC(CCNCCOC1=CC=C2C=CC=CC=C2C=C1)O.Cl

デキサメタゾン

392.45

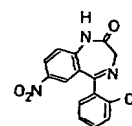
CC(C)[C@H](O)[C@@H](OC(=O)CO)[C@H](O)[C@@H](C)[C@H]1CC[C@@H]2[C@@]1(CC[C@H]3[C@H]2CC=C4[C@@]3(CC[C@@H](C4)O)C)C

【圖 6-2】

化合物
クロリゼン

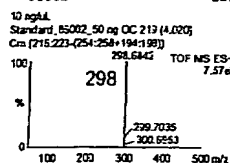
MW
315.72

構造式



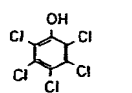
86002

297.4

C1CN(C1)C(C2=CC=CC=C2F)C3=CC=CC=C3N

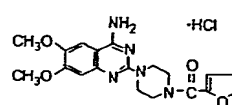
ベンタクロロフェノール 226.34

226.34

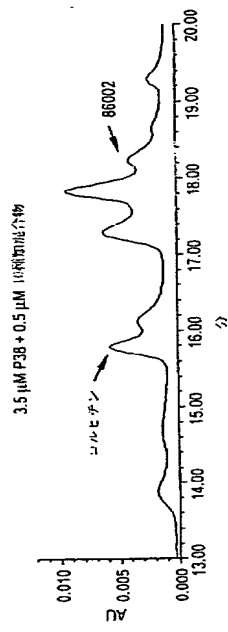
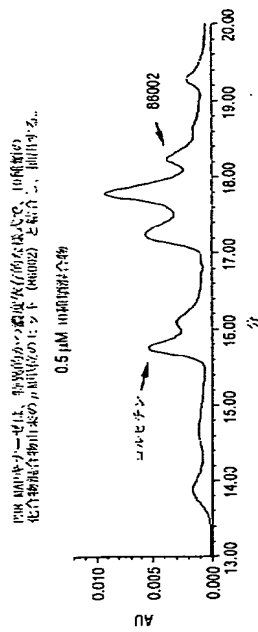


ノラノシン

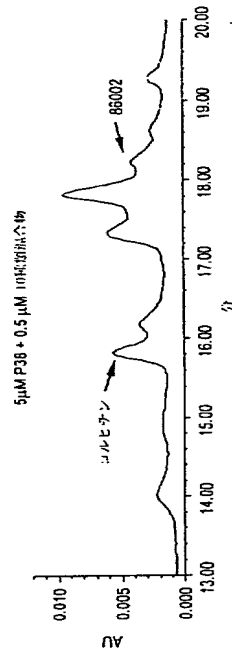
419.9



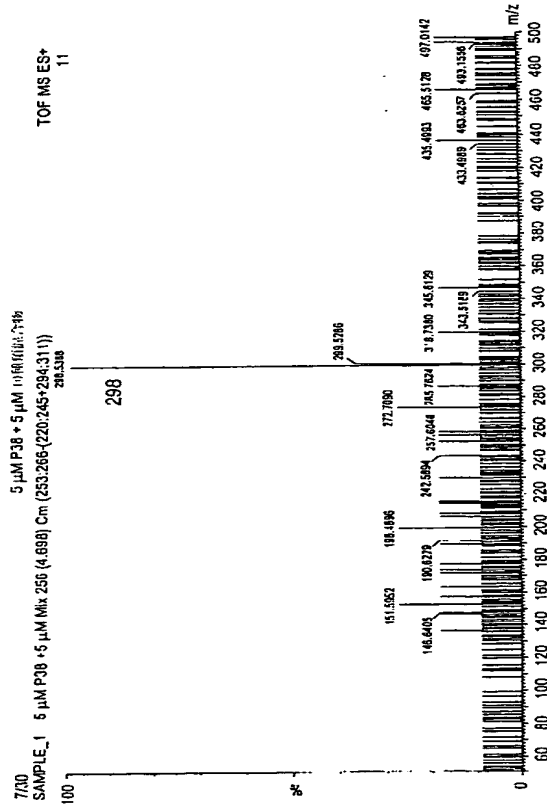
【図 7-1】



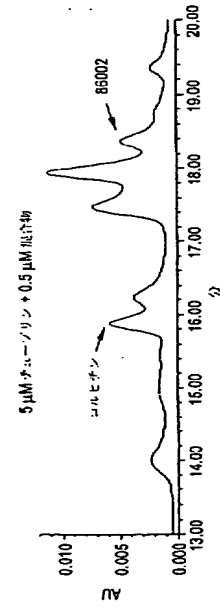
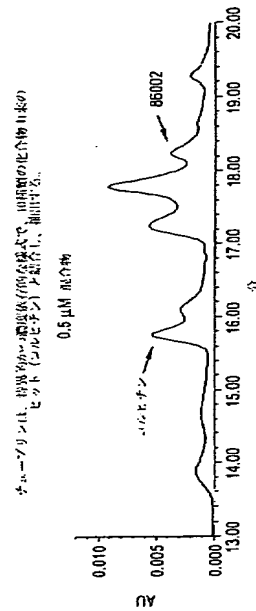
【図 7-2】



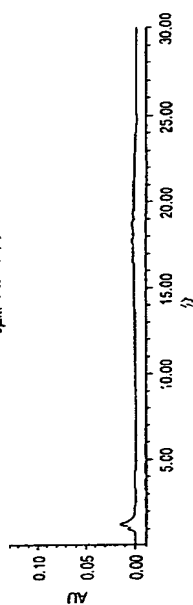
【図 7-3】



【図 8-1】



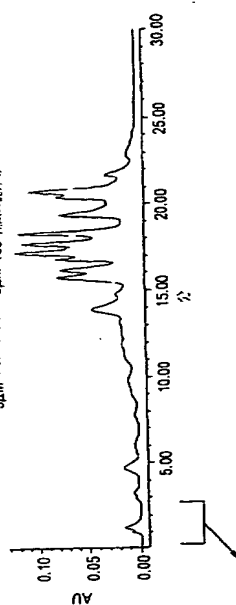
チ、アゾノール、特許のカラミッド依存的な組成で、100%以上の化合物混合物による。ピクト（ピクトサン）と結合し、抽出する。



21M F38 + 20μM 100 cmol mix-100nM B6002
microalgae 13 TM-01-476 218 (4.019) Cm (215:223(277:288+151:167))

Mass spectrum showing relative intensity (%) versus m/z. The base peak is at m/z 298. Other significant peaks are labeled with their m/z values.

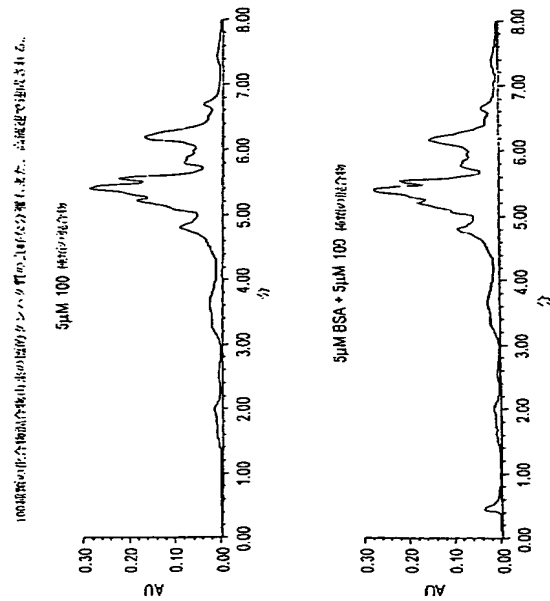
m/z	Relative Intensity (%)
18.0051	~1
102.0972	~1
163.8139	~10
177.7820	~5
191.7859	~15
215.8343	~10
218.8000	~10
232.8832	~5
254.8018	~10
271.8602	~10
312.8440	~10
333.3118	~5
358.5472	~5
370.7198	~5
373.6647	~5
413.3710	~5
415.5208	~5
424.5673	~5
433.8894	~5
448.5418	~5
477.8350	~5
298	100

 $5\mu\text{M}$, γ -irradiation dose = 100 kGy 

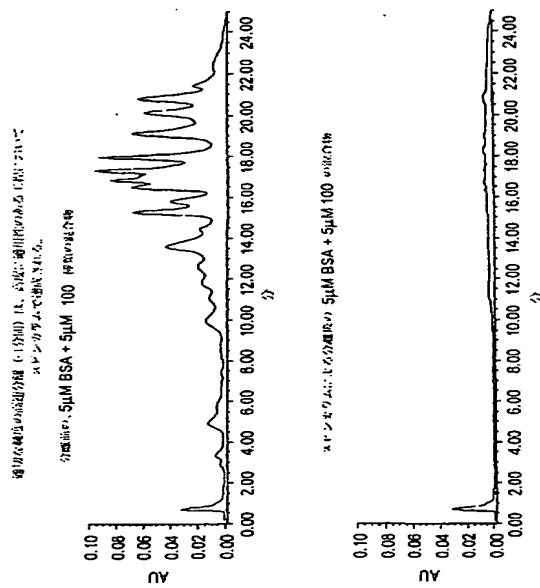
【図 1 1 - 3】



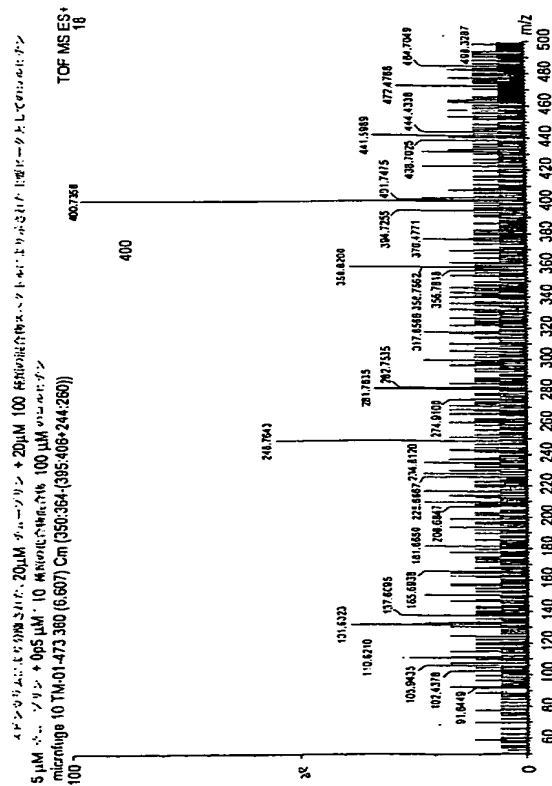
【図 1 2】



【図 1 3 - 1】



【図 1 3 - 2】



【国際公開パンフレット】

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100

(43) International Publication Date
1 August 2002 (01.08.2002)

PCT

(10) International Publication Number
WO 02/058533 A2

- (51) International Patent Classification: A61R (74) Agent: CT.ARN, Paul, T., Clark & (Filing 11 P, 126 Federal Street, Boston, MA 02110-2214 (US).
- (21) International Application Number: PCT/US01/43348 (81) Designated States (regionally): AE, AG, AI, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, FR, GB, GR, GT, HK, HN, HU, ID, IL, IN, IS, JP, KE, KP, KR, KZ, LC, LK, LR, LS, LU, LV, MA, MD, MG, MK, MN, MW, MX, MY, ND, NZ, PL, PT, RU, SI, SR, SG, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.
- (22) International Filing Date: 19 November 2001 (19.11.2001) (84) Designated States (regionally): ARIPU patent (GL, GM, IL, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), European patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IT, LI, NL, PT, SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GN, GW, ML, MR, NI, SN, TD, TG).
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data: 60/249,322 17 November 2000 (17.11.2000) US
60/259,451 15 October 2001 (15.10.2001) US
- (43) Related by continuation (CON) or continuation-in-part (CIP) to earlier applications: 60/249,322 (CIP) US
Filed on 17 November 2000 (17.11.2000)
60/259,453 (CIP) US
Filed on 15 October 2001 (15.10.2001)
- (71) Applicant and
(72) Inventor: SLANETZ, Alfred, E. [US]; 14 Nichols Road, Cohasset, MA 02025 (US).
- Published:
without international search report and to be republished upon receipt of that report
- For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 02/058533 A2

(54) Title: PROCESS FOR DETERMINING TARGET FUNCTION AND IDENTIFYING DRUG LEADS

(57) Abstract: The present invention relates to methods for using chemical ligands to determine target function and identify drug leads.

WO 02/058533

PCT/US01/43348

5 PROCESS FOR DETERMINING TARGET FUNCTION AND
 IDENTIFYING DRUG LEADS

Background of the Invention

10 1. INTRODUCTION

 The present invention relates to a method of exposing targets to a plurality of potential ligands, collecting ligand—target pairs, using the ligand to analyze the target's biological function, and optionally identifying the ligand chemically and/or structurally. In one embodiment of the invention ligands are selected which bind to pharmaceutically relevant targets. In another
15 embodiment of the invention, ligand—target pairs are collected and analyzed on a genomic scale. The invention further relates to a method of screening a plurality of potential ligands in at least one bioassay for a change in phenotype and using the hit(s) to identify the corresponding molecular target.

20

 2. BACKGROUND OF THE INVENTION

 2.1. TRADITIONAL APPROACH TO DRUG DISCOVERY

 In general drugs discovered in the last 50 years are based on a few hundred targets and there are presently about 450 validated targets used for
25 screening by all of the pharmaceutical companies combined. These targets have typically been developed using the traditional approach to drug discovery in which the target is validated using reductionist biology including gene over-expression, gene knockout, gene sequence homology searching for functional domains, x-ray crystallography,

30

WO 02/058533

PCT/US01/43348

or specific cellular and biological assays. Furthermore in drug discovery as it is practiced today, target validation, assay development, high throughput screening and lead generation are performed in series.

2.2. GENOMICS

5 The large number of uncharacterized genes from the completion of the sequencing of the human genome makes it difficult but essential for a pharmaceutical company to validate and choose only the right target to unleash the value of the human genome sequence. It is estimated that of the 100,000 or more genes in the human genome, at most 10,000 of these genes will be
10 pharmaceutically useful targets. This huge number of genes is overwhelming the reductionist approach to gene validation thereby presenting a major bottleneck in drug discovery.

The accumulating mass of DNA sequence data has given rise to the field of functional genomics that promises to alleviate the bottleneck. Gene
15 expression profiling can be studied using DNA arrays (De Risi JL *et al.*, 1997, *Science* 278 :680). Protein expression profiling can be performed using protein arrays (Pawelczak CP *et al.*, 2000, *Drug Dev. Research* 49:34). Gene function can be studied by the introduction or mutation of a gene to induce a conditional change in phenotype. Alternatively, an antisense or ribozyme version of a gene
20 may be expressed in a variety of cell lines or organisms including transgenic or knockout mice, *C. elegans*, zebra fish, *Drosophila* or yeast (Conture LA *et al.*, 1996, *Trends in Genetics* 12:510; Nadeau JH *et al.*, 1998, *Curr. Opin. Genet. Dev.* 8, 311).

Differential gene expression can be detected using a variety of techniques
25 including: differential screening (Todder TP *et al.* 1988 *PNAS* 85:208), subtractive hybridization (Hedrick SM *et al.* 1984, *Nature* 308:149), differential display (Jiang P and Pardoll A 1993 *TISS*262311), gene microarray (Lockhart, D *et al.*, 1996, *Nature Biotechnology* 14:1675; Schena M *et al.*, 1995, *Science* 270: 467; 2000, *Nature Genetics* 24:236), representational difference analysis
30 (Hubank M *et al.*, 1994, *Nucleic Acids Research* 22:5640), large scale sequencing of expressed sequence tags (ESTs), reverse transcriptase PCR, serial analysis of gene expression (SAGE; Nacht M *et al.*, 1999, *Cancer Res.* 59:5464)

WO 02058533

PCT/US01/43348

and laser capture microdissection (Sgami DC *et al.*, 1999, Cancer Research 59:5656). Microarray technology represents the current state of the art for genomics and has been used to study cell cycles, biochemical pathways, genome wide expression in yeast, cell growth, cell differentiation, cell responses to a single compound, genetic diseases (M. Schena, 1998, TIBTECH 16:301).

2.3. IDENTIFICATION AND CHARACTERIZATION OF PROTEIN TARGETS

Using classical biochemical techniques, previously unknown receptors for small molecules have been identified at the protein level using *in vitro* biochemical methods including photo-crosslinking, radiolabeled ligand binding and affinity chromatography (Jakoby WB *et al.*, 1974, Methods in Enzymology 46:1). These methods require purification of the protein. In order to clone the gene for the receptor, the peptide must be further sequenced and this sequence used to clone the cDNA for the protein. Small molecules can be radiolabeled and used to determine the molecular target (Kwon HJ *et al.*, 1998, PNAS 95:3356). Alternatively, small molecules can be immobilized on an agarose matrix and used to screen extracts of a variety of cell types and organisms. For example, purvalanol B (a known inhibitor of cyclin-dependent kinases) was immobilized on an agarose matrix and used to screen extracts from a diverse collection of cell types and organisms and a number of proteins with kinase activity were isolated (Knockaert M *et al.*, 2000, Chem. Biol. 7:411). Alternatively, trapoxin is a cyclotetrapeptide that inhibits histone deacetylation and arrests the cell cycle. Two nuclear proteins co-purified with histone deacetylase activity from fractionated cell extracts on an affinity matrix covalently modified with trapoxin. Subsequently the proteins were sequenced and cDNAs encoding the proteins were cloned from a cDNA library (Tantton J *et al.*, 1996, Science 272:408).

Currently, the primary system for studying protein-protein interactions is the yeast two hybrid system. In this approach, one protein is fused to the DNA binding domain and another protein is bound to the DNA activation domain of a eukaryotic transcription factor and expressed in the presence of a reporter gene

WO 02/058533

PCT/US91/43348

which allows the yeast to grow. If the two heterologous proteins bring the two domains together, then the yeast containing the proteins which interact are selected by growth (Fields S *et al.*, 1989, Nature 340:245).

A yeast "three hybrid" transcription activation system has been used to
 5 clone a gene encoding a previously identified receptor for the drug FK506. This three hybrid system displays an anchored derivative of the active ligand against a library of cDNAs fused to the transcriptional activation domain (Borchardt A. *et al.*, 1997, Chem. Biol. 4:961; Licita EJ *et al.*, 1996, PNAS 93:12817). In Licita *et al.*, the hormone binding domain of the rat glucocorticoid receptor was fused to the Lex A DNA binding domain, a cDNA encoding the FK506 receptor (FKBP12)
 10 (FKBP12) was fused to the transcriptional activation domain and the two were expressed in the yeast two hybrid system. The yeast cells were plated on medium containing a heterodimer of covalently linked dexamethasone and FK506 and the cells grew in a way that may be inhibited by nonimerized FK506. When the experiment was repeated with a cDNA expression library fused to the transcriptional activation domain in place of the cDNA encoding FK506 binding protein, the yeast which grew contained cDNA clones encoding the FK506 binding protein. However, this experiment was done using a chemical interacting with an known target. In Borchardt A *et al.*, yeast cells in the presence of a FKBP12-GAL4 DNA binding domain fusion, the FR domain of the FK506 binding protein rapamycin associated protein, and rapamycin transcribe the HIS3 3 reporter genes allowing the cells to grow in the absence of histidine (Borchardt A *et al.*, 1997, Chem Biol 4:961).

Expression cloning can be used to test for the target within a small pool of proteins (King RW *et al.*, 1997, Science 277:973). Peptides (Kieffer *et al.*, 1992, PNAS 89:12048), nucleoside derivatives (Hanstalter KA *et al.*, 1999, Curr. Biol. 9:174), and drug-bovine serum albumin (drug-BSA) conjugate (Tanaka *et al.*, 1999, Mol. Pharmacol. 55:356) have been used in expression cloning.

33 Another useful technique to closely associate ligand binding with DNA encoding the target is phage display. In phage display, which has been predominantly used in the monoclonal antibody field, peptide or protein libraries

WO 02/058533

PCT/US91/43348

are created on the viral surface and screened for activity (Smith GP, 1985, Science 228:13:5). Phage are panned for the target which is connected to a solid phase (Parmley SF *et al.*, 1988, Gene 73:305). One of the advantages of phage display is that the cDNA is in the phage and thus no separate cloning step is required. Dyax has used a phage display affinity column to isolate macromolecules but not small molecules (US97/04425).

Recently, Sche *et al.* used the natural product FK506 as an affinity probe to clone FKBP12 from a T7 cDNA phage display library. They used an affinity matrix bearing biotinylated FK506 to screen a phage library prepared with human brain cDNA. The phage particles remaining after two rounds of affinity selection shared a common 450 bp insert which corresponded to full length FKBP12.

Alternatives to phage display include plasmid display (Cull *et al.*, 1992, PNAS 89:1865; Schatz PJ *et al.*, 1996, Methods Enzymol 267:171), polysome display (Matheakis LC *et al.*, 1996, PNAS 91:9022; Matheakis LC, 1996, Methods Enzymol 267:195), protein tagging (Whitehorn EA *et al.*, 1995, Biotechnology 13:1215), ribosome display (Hanes J *et al.*, 1998, PNAS 95:14130), and cell surface display in bacteria and eukaryotes (Georgiou G *et al.*, 1997, Nat. Biotechnol 15:29; Chesnut J *et al.*, 1996, J. Immun Methods 193:17). Peptides or proteins can also be linked chemically via paratubercin to the mRNA that encodes it (Roberts R *et al.*, 1997, PNAS 94: 12297).

2.4. CHEMICAL GENETICS

Chemical genetics is a new and potentially powerful approach to defining gene function through the use of chemicals to cause a conditional change in gene expression or gene function. However, to date, it has not advanced far from traditional drug discovery using traditional high throughput cell based screening assays against known targets to which drugs are already available to find more hits to those targets. The current status of chemical genetics is demonstrated in the work of Haggarty SJ *et al.* (2000, Chem Biol 7:275) in which 139 compounds were identified from a high throughput screen of the Chembridge Diverset library for inhibition of mitosis in a cell based assay and then assayed in

WO 02/058533

PCT/US99/43348

an *in vitro* tubulin polymerization assay. Of the 139 compounds, 52 were
antagonists which destabilized tubulin by the same mechanism as colchicines.
One compound was demonstrated to be an agonist which stabilized tubulin by
the same mechanism as taxol. 86 compounds had no effect and thus likely
5 modulated mitosis via non-tubulin targets. For the compounds targeting
non-tubulin targets based upon visible effects on the chromosomes and
cytoskeleton, 7 were believed to be weak antagonists of tubulin and one
(monastrol) was demonstrated to inhibit the kinesin-related protein Eg5 (Mayer
et al., 1999, Science 286:971). In the case of Haggarty SJ et al., low affinity
10 ligands were selected since assays were performed using a ligand concentration
of 20 to 50 μ M. However, low affinity ligands are of limited value in
determining target function.

Rossini GR et al. identified a novel small molecule, myoseverin, by a
cell morphological screen which binds to tubulin to induce the reversible fission
and proliferation of muscle cells. Unlike the current invention, Schulz is relying
15 on the standard functional genomics DNA array approach to understand the
mechanism (Rossini GR et al., 2000, Nat Biotechnol 18:304). Chemicals have
been used to study function since colchicines were shown to have an effect on
mitosis in 1889
20 (Eiselt O, 1949, Science 110:692). However, current practice is limited to
identifying ligands which bind to known targets or to unidentified targets which
result in a particular phenotype.

Previous efforts to characterize the function of unknown genes are
exemplified by orphan receptor analysis. Orphan receptors are encoded by
25 genes which share DNA sequence similarity with previously identified receptors.
On that basis, such sequences are placed into a receptor superfamily for which
the natural physiological role and ligand are unknown. The present state of the
art is to use genetic techniques or to use drugs or protein ligands known to bind
to other members of the family to determine their function (Werne M et al.,
30 2000, Brain Res 863:112; Dordji K. et al., 2000, J. Biol. Chem. 275:12243;
Yang C., 1999, Cancer Res. 59:4519; Chion L, 1999, Br. J. Pharmacol 128:103;
Williams C, 2000, Curr. Opinion in Biotechnology 11:42).

WO 02/058533

PCT/US01/43348

2.5. CHEMICAL TARGET CHARACTERIZATION

Once a target is validated, two major screening categories are applied:

bioassays and mechanism based assays (Gordon et. al., 1994, J. Med. Chem. 37:1386). Bioassays measure an effect on a cell of the compounds being screened on viability or metabolism. For example, penicillin was discovered by its growth inhibition in bacterial culture. Mechanism based assays include biochemical assays measuring an effect on enzymatic activity, cell based assays in which the target and a reporter system (e.g., luciferase or β -galactosidase) have been introduced into a cell (Mouks A et. al., 1997, Anticancer Drug Des. 12: 523), or binding assays. Binding assays can be performed with the target fixed to a well, bead (Boerwuth N et al., 1989, Nature 1989, 341:167; Meldal M, 1994, PNAS 91, 3314) or chip (Sunberg S, 2000, Curr. Opin. In Biotechnol 11:47) or captured by an immobilized antibody, and the bound ligands are detected usually using calorimeter or by measuring fluorescence (Sunberg S, 2000, Curr. Opin. In Biotechnology 11:47).

In some newer binding assays, molecules binding to a target of known function have also been resolved by capillary electrophoresis (US 5783397; US99/15458). In other new assays, libraries were weight-coded and deconvoluted using mass spectroscopy (Carell T et al., 1995, Chem Biol. 2: 171; Fang AS et. al., 1998, Comb Chem High Throughput Screen 1:23; US 99/23837; US99/00024). HPLC has also been used with mass spectroscopy to characterize combinatorial library purity and to analyze metabolites in plasma samples (Korfmacher WA et al., 1999, Rapid Commun Mass Spectrom 13:1991; Zeng L et al., 1998; Comb Chem High Throughput Screen 1:101; Nedved ML et al., 1996, Anal Chem 68: 4228; Zimmer D et al., 1999, J. Chromatogr A 854:23; Aubagnac JL, Comb Chem High Throughput Screen 2:289).

3. SUMMARY OF THE INVENTION

The present invention relates to the use of a target of unknown function to select for small molecules from a chemical library which are then used in an assay to determine the target's function. According to the invention, members of

WO 02/058533

PCT/US01/43348

the chemical library are mixed with the protein in a biochemical binding assay and those that bind are then (sequentially or in parallel) used in a *in vitro* or *in vivo* bioassay to determine the function of the gene by a change in a measurable phenotype in a biological or pathological condition.

5 Alternatively, the invention uses chemicals which induce a phenotypic change in a bioassay to determine the identity of the target. The invention provides a method of screening a plurality of potential ligands in at least one bioassay, selecting ligands which produce a change in phenotype in a bioassay, and using the ligand to screen candidate targets to identify the particular target(s) responsible for the altered phenotype.

10 The invention can be used to define the function of genes and to simultaneously validate the drug target and generate a drug lead thus streamlining the drug discovery process. The structure activity relationship information provided by the parallel comparison of a large number of structurally diverse hits which bind to the target but have different activities in phenotypic assays can be used to rapidly optimize the lead. Using the invention, 15 the massive numbers of genes provided by genomics can be systematically sorted and useful drug targets can be validated and selected for a given disease.

The present invention is different from the art because the latter describes 20 screening against a known target while the present invention does not require any prior knowledge of target identity or function. Furthermore, the present invention does not absolutely require the constraint of a predetermined subunit of a particular mass in the construction of its library. According to the invention, virtually any ligand library produced by combinatorial or noncombinatorial means may be used. Non-limiting examples include chemical, peptide, natural 25 product, natural product-like, sugar or antibody libraries. Peptides and proteins can be made to cross the cell membrane using a sequence from HIV TAT, HSV VP22 or Antennapedia peptides containing protein transduction domains (Swartz SR *et al.*, 2000, Trends in Cell Biology 10:290). Libraries may consist of pools of ligands or may be collections of single ligands screened individually.

30 Accordingly, in one aspect, the invention features a method for selecting a candidate ligand which binds a target molecule. This method involves

WO 01/058533

PCT/US01/43348

contacting an *in vitro* sample including a target molecule with a library of candidate ligands under conditions that allow complex formation between the target molecule and one or more of the candidate ligands. The complex is isolated, and one or more of the candidate ligands are recovered from the complex. Additionally, one or more recovered candidate ligands are identified.

In various embodiments of the above aspect, the target molecule is a molecule of unknown biological function or a molecule that has not been previously validated as a drug target. In other embodiments, the library includes at least two different chemical scaffolds or includes at least 11 different compounds. In other embodiments, the complex is isolated using size exclusion or biphasic chromatography (e.g., chromatography using an internal surface reverse phase (ISRP), GFF, or GFFII resin). In other embodiments, MS, IR, FTIR, NMR, and/or UV analysis is used to identify the recovered candidate ligand. In yet other embodiments, the method includes determining the mass to charge ratio of a parent peak, a fragment peak, and/or an isotope peak in the mass spectrum of the recovered candidate ligand. In one embodiment, the method also includes contacting the sample with a competitor ligand known to bind the target molecule. This competitor may reduce the number of low affinity candidate ligands that bind the target molecule, allowing the higher affinity candidate ligands to be selected.

In another aspect, the invention features another method for selecting a candidate ligand which binds a target molecule. This method involves contacting an *in vitro* sample including a first target molecule and a second target molecule with a library of candidate ligands under conditions that allow complex formation between the first target molecule and one or more of the candidate ligands and allow complex formation between the second target molecule and one or more of the candidate ligands. A first complex including the first target molecule bound to a candidate ligand and a second complex including the second target molecule bound to a candidate ligand are isolated. One or more of the candidate ligands from the first complex and/or from the second complex are recovered and identified. In one embodiment, the method

WO 02/058533

PCT/US01/43348

also includes contacting the sample with a competitor ligand known to bind the first target molecule or the second target molecule.

Additionally, the invention provides various methods for determining the biological function of a target molecule, such as a naturally or non-naturally occurring protein, nucleic acid, carbohydrate, or other organic molecule. The methods may be used to determine the function of a gene or a protein of interest, such as gene or protein that is upregulated or downregulated in a particular disease state or in the presence of a particular biological stimuli (such as TNF α). The methods may also be used to identify therapeutically active compounds for the treatment of a disease state.

In one such aspect, the invention provides a method for determining the biological function of a target molecule. This method includes contacting an *in vitro* sample including a target molecule with a library of candidate ligands under conditions that allow one or more of the candidate ligands to bind the target molecule. A candidate ligand which binds the target molecule is selected. The effect of the selected candidate ligand in a biological assay is measured, thereby determining the biological function of the target molecule. In various embodiments, target molecule is a molecule of unknown biological function or a molecule that has not been previously validated as a drug target. In other embodiments, the target molecule is upregulated or downregulated in a disease state, in the presence of a physiological stimulus (e.g., a cytokine such as TNF), or during a specific cellular or biological process. In particular embodiments, the target molecule is upregulated or downregulated during angiogenesis, differentiation, proliferation, or insulin secretion. In one embodiment, the selected candidate ligand is identified using a method such as MS, IR, FTIR, NMR, UV, or any other appropriate method. In particular embodiments, the selected candidate ligand increases the activity of the target molecule in the biological assay. For example, the candidate ligand may activate an activity of the target molecule (such as an enzymatic activity), promote the production of the target molecule, increase the stability of the target molecule, alter the localization of the target molecule, or promote the association of the target molecule with another molecule. In other embodiments, the selected candidate

WO 02/058533

PCT/US01/43348

ligand decreases the activity of the target molecule in the biological assay. For example, the candidate ligand may inhibit an activity of the target molecule, inhibit the production of the target molecule, decrease the stability of the target molecule, alter the localization of the target molecule, or inhibit the association of the target molecule with another molecule. Exemplary biological assays include a throughput screen using a nontransfected cell line, cell, tissue, or other biological system where the target is not previously known. In other embodiments, the biological assay involves determining the effect of the selected candidate ligand on a tissue from an organism having a disease or disorder or undergoing a specific cellular or biological process in the presence or absence of a physiological stimulus is measured, thereby determining the biological function of the target molecule. In one embodiment, the tissue is a mammalian tissue, such as a human tissue.

Methods for crosslinking two ligands with bind the same target molecule are also provided. These methods allow one or more target surfaces to promote or catalyze the reaction between two ligands. These methods may be used to screen a library of ligands to determine what ligands bind the target molecule and what crosslinked products containing a combination of ligands bind the target molecule with the highest affinity. The crosslinked products may be used as lead compounds in the development of therapeutics or used to characterize the active site of the target molecule. Related methods may be used to crosslink two ligands with bind different target molecule. These methods may be used to determine what target molecules interact with a target molecule of interest, thereby determining what molecules are in the same pathway as the target molecule of interest.

In another aspect, the invention features a method for reacting two ligands that bind a target molecule of interest. This method involves contacting a cell or *in vitro* sample including a target molecule with a first ligand (e.g., a first ligand having a first crosslinker) and with a second ligand under conditions that allow the target molecule to bind both the first ligand and the second ligand and allow the first crosslinker to covalently bind the second ligand, thereby generating a crosslinked product including the first ligand and the second ligand.

WO 02/053533

PCT/US01/43348

In some embodiments, target molecule is a molecule of unknown secondary or tertiary structure. In other embodiments, the location or the tertiary structure of the binding site in the target molecule for the first ligand or the second ligand is unknown. In a particular embodiment, the affinity of the crosslinked product for the target molecule is greater than the affinity of the first ligand or the second ligand for the target molecule. In another embodiment, the crosslinked product is used for drug discovery or development, lead optimization, or development of an agricultural or environmental agent. In yet another embodiment, the target molecule promotes or catalyzes the reaction between the first and second ligands. In another embodiment, the first ligand is reacted with a crosslinker prior to being contacted with the target molecule. In yet another embodiment, the first ligand, the second ligand, and a crosslinker are reacted in the presence or absence of the target molecule.

In another aspect, the invention features a method for reacting two ligands that bind different target molecules. This method includes contacting a cell or *in vitro* sample including a first target molecule and a second target molecule with a first ligand (e.g., a first ligand having a first crosslinker) and with a second ligand. The contacting is conducted under conditions that allow (i) the first target molecule to bind the first ligand, (ii) the second target molecule to bind the second ligand, and (iii) the first crosslinker to covalently bind the second ligand, thereby generating a crosslinked product including the first ligand and the second ligand. In one embodiment, the location or the tertiary structure of the binding site in the first target molecule for the first ligand and/or the location or the tertiary structure of the binding site in the second target molecule for the second ligand is unknown. In one embodiment, the generation of the crosslinked product indicates that the first target molecule (e.g., a protein) and the second target molecule (e.g., a protein) interact *in vivo* or are part of the same biological pathway. In another embodiment, the crosslinked product is used for drug discovery or development, lead optimization, or development of an agricultural or environmental agent. In yet another embodiment, one or both target molecules promote or catalyze the reaction between the first and second ligands. In another embodiment, the first ligand is reacted with a crosslinker

WO 02/058533

PCT/US01/43348

prior to being contacted with the target molecules. In yet another embodiment, the first ligand, the second ligand, and a crosslinker are reacted in the presence or absence of the target molecules.

In another aspect, the invention provides a method for isolating a second protein which binds a first protein. This method involves contacting a cell or an *in vivo* sample including a first protein and a second protein with a first ligand having a first crosslinker and with a second ligand. The contacting is conducted under conditions that allow (i) the first protein to bind the first ligand, (ii) the second protein to bind the second ligand, and (iii) the first crosslinker to covalently bind the second ligand, thereby generating a crosslinked product including the first ligand and the second ligand and generating a complex including the crosslinked product, the first protein, and the second protein. The complex is isolated, and the first protein and/or the second protein in the complex or recovered from the complex is identified. In one embodiment, the first and/or second protein includes a detectable group. In another embodiment, the second ligand includes a crosslinker. In one embodiment, the generation of the crosslinked product indicates that the first protein and the second protein interact *in vivo* or are part of the same biological pathway. In another embodiment, the crosslinked product is used for drug discovery or development, lead optimization, or development of an agricultural or environmental agent.

The invention also provides numerous methods for selecting a target molecule which binds a compound of interest. For example, the compound may be a molecule that appears to promote or inhibit a disease state. The selected target molecule may be used, for example, to study the disease, to identify other molecules associated with the disease, and to identify therapeutics which bind or modulate the activity of the target molecule or another member of the disease pathway.

In another aspect, the invention provides a method for selecting a candidate target molecule which binds a small molecule of interest. The method involves contacting an *in vivo* sample including a small molecule of interest with a library of candidate target molecules under conditions that allow complex formation between the small molecule of interest and one or more of the

WO 02/058533

PCT/US01/43348

candidate target molecules. The complex is isolated, and one or more of the candidate target molecules are recovered from the complex, thereby selecting one or more candidate target molecules which bind the small molecule of interest. In various embodiments, the library of candidate target molecules is
5 recombinantly produced or is obtained from an extract from a cell, tissue, or organism. The library of candidate target molecules can be unpurified, partially purified, or completely purified from other components prior to being contacted with the small molecule of interest. In various embodiments, the target molecules are expressed on the surface of phage or are not expressed on the
10 surface of phage. In one embodiment, prior to contacting the small molecule with the library of candidate target molecules, the small molecule of interest is selected from a library of small molecules based on its effect in a biological assay. In one embodiment, the method also includes identifying the selected target protein. In particular embodiments, the small molecule of interest has a
15 moiety other than an amino acid or has a molecular weight less than 5000, 4000, 3000, 2000, 1000, 750, 500, or 250 daltons.

In another aspect, the invention provides a method for selecting a target protein which binds a small molecule of interest. This method includes expressing in a population of cells a protein fusion including a target protein
20 covalently linked to surface protein, the expression being carried out under conditions that allow the display of the protein fusion on the surface of the cells. The cells are contacted with a small molecule of interest, and the cells which bind the small molecule of interest are selected, thereby selecting the target proteins which bind the small molecule of interest. Exemplary cells include
25 mammalian, bacterial, yeast, and insect cells. In one embodiment, the method also includes identifying the selected target protein. In particular embodiments, the small molecule of interest has a moiety other than an amino acid or has a molecular weight less than 5000, 4000, 3000, 2000, 1000, 750, 500, or 250 daltons.

30 In another aspect, the invention features another method for selecting a target protein which binds a small molecule of interest. This method involves expressing in a population of cells a protein fusion including a target protein

WO 02/058533

PCT/US01/43348

covalently linked to surface protein, the expression being carried out under conditions that allow the display of the protein fusion on the surface of viruses released from the cells infected with the virus. The viruses are contacted with a small molecule of interest, and the viruses which bind the small molecule of interest are selected, thereby selecting the target proteins which bind the small molecule of interest. In one embodiment, the method also includes identifying the selected target protein. In various embodiments, the virus is a bacteriophage or adenovirus. In particular embodiments, the small molecule of interest has a moiety other than an amino acid or has a molecular weight less than 5000, 4000, 3000, 2000, 1000, 750, 500, or 250 daltons. In yet other embodiments, the small molecule of interest does not contain biotin or is not naturally produced by bacteria. In still other embodiments, the small molecule of interest is a nucleic acid, lipid, or carbohydrate. In still other embodiments, the small molecule of interest is immobilized on a solid surface such as a magnetic or fluorescent bead. In other embodiments, an adenovirus is used to infect 293 cells or perov cells, or a bacteriophage is used to infect bacteria.

In another aspect, the invention features a method for selecting a target protein which binds a small molecule of interest. This method involves expressing in a population of cells or an *in vitro* sample a library of target proteins in which each target protein is covalently linked to a nucleic acid encoding the target protein. The cells or *in vitro* sample are contacted with a small molecule of interest, and the target proteins which bind the small molecule of interest are selected. In one embodiment, the method also includes identifying the selected target protein. In particular embodiments, the small molecule of interest has a moiety other than an amino acid or has a molecular weight less than 5000, 4000, 3000, 2000, 1000, 750, 500, or 250 daltons.

In various embodiments of any of the above methods for selecting a target molecule or target molecule which binds a small molecule of interest, at least 2, 5, 10, 20, 50, 100, 1000, 10000, or more target molecules are contacted with the small molecule. In other embodiments, a target peptide or protein is associated with a polynucleotide encoding the target, using standard methods such as phage display, cell surface display, plasmid display, ribosome display,

WO 02/058533

PCT/US01/43348

viral display). In other embodiments, the small molecule is immobilized on a solid surface, such as a column, bead, or magnetic bead. In other embodiments, the small molecule contains a fluorescent group, or the small molecule is indirectly or directly linked to a fluorescent group (e.g., linked through the binding of a fluorescently labeled antibody), and the complex of the small molecule and a target molecule is isolated using FACS sorting. In other embodiments, the small molecule of interest is a non-naturally occurring molecule or a naturally occurring molecule from an organism other than bacteria (e.g., such as a naturally occurring human molecule).

The invention also provides methods for identifying compounds that bind a target molecule before the target molecule is experimentally validated as a drug target. Additionally, methods are provided for identifying ligands for two or more target molecules. For example, binders can be simultaneously identified for multiple target molecules by performing an assay containing multiple target molecules or by performing multiple assays in parallel. These high throughput assays greatly increase the number of target molecules that can be analyzed.

Accordingly, in one aspect, the invention provides a method for selecting a candidate compound that binds or modulates the activity of a target molecule prior to validation of the target molecule as a drug target. This method involves contacting a cell or an *in vitro* sample including a target molecule that has not been previously validated as a drug target with a library of candidate compounds under conditions that allow one or more of the candidate compounds to bind or modulate the activity of the target molecule. A candidate compound which binds or modulates the activity of the target molecule is selected. In one embodiment, the selected candidate compound is identified. In other embodiments, the method also includes measuring the effect of the selected candidate compound in a biological assay, thereby determining the biological function of the target molecule. In yet other embodiments, the cell or *in vitro* sample includes at least 2, 5, 10, 20, 30, 50, 100, or more target molecules, and for each of the target molecules, a candidate compound is selected that binds or modulates the activity of the target molecule.

WO 02/058533

PCT/US01/43348

In another aspect, the invention features a method for selecting candidate compounds that bind or modulate the activity of target molecules. This method involves contacting a cell or an *in vitro* sample including a first target molecule and a second target molecule with a library of candidate compounds under conditions that allow one or more of the candidate compound to bind or modulate the activity of the first target molecule and allow one or more of the candidate compound to bind or modulate the activity of the second target molecule. A candidate compound which binds or modulates the activity of the first target molecule is selected, and a candidate compound which binds or modulates the activity of the second target molecule is selected. In one embodiment, one or more of the selected candidate compounds are identified. In other embodiments, the method also includes measuring the effect of one or more of the selected candidate compounds in a biological assay, thereby determining the biological function of the target molecule. In yet other embodiments, the cell or *in vitro* sample includes at least 5, 10, 20, 30, 50, 100, or more target molecules, and for each of the target molecules, a candidate compound is selected that binds or modulates the activity of the target molecule.

The invention also features a variety of databases. These databases are useful for storing the information obtained in any of the methods of the invention. These databases may also be used in the development of therapeutics and in the selection of a preferred therapeutic for a particular patient or class of patients. Many other uses of these databases are described herein.

In one such aspect, the invention features an electronic database including at least $10, 10^2, 10^3, 10^4, 10^5, 10^6, 10^7, 10^8$, or 10^9 records of target molecules correlated to records of ligands and their ability to bind or modulate the activity of the target molecules. In a related aspect, the invention provides an electronic database including a plurality of records of target molecules that have not been previously validated as drug targets and/or target molecules of unknown biological function correlated to records of ligands and their ability to bind or modulate the activity of the target molecules. In another related aspect, the invention features an electronic database including at least $10, 10^2, 10^3, 10^4, 10^5, 10^6, 10^7, 10^8$, or 10^9 records of target molecule domains correlated to records

WO 02/058533

PCT/US01/43348

of ligands and their ability to bind the domains. By "domain" is meant a domain found in one or more proteins that catalyze the same type of reaction or that bind the same type of molecules; or the domains are identified as different protein structural motifs or functional families based upon the analysis of DNA or amino acid sequences, x ray crystal structures, or biological assays. For example, the database may contain records of ligands and their ability to bind a kinase domain (i.e., able to bind one or more kinases) or a phosphatase domain (i.e., able to bind one or more phosphatases). This database may be used, for example, for characterizing the binding sites of proteins or other target molecules and for determining the selectivity of ligands for particular binding sites or particular families of compounds.

In various embodiments of the above databases, the database includes records for at least 0.5, 1, 5, 10, 20, 30, 40, 50, 60, 70, 80, 90, or 100% of the proteins or protein domains in the proteome of an organism, such as a bacteria, yeast, or mammal. In particular embodiments, the database includes records for at least 0.5, 1, 5, 10, 20, 30, 40, 50, 60, 70, 80, 90, or 100% of the proteins or protein domains in the human proteome. In yet other embodiments, the database includes records for at least one protein expressed by an open reading frame for at least 0.5, 1, 5, 10, 20, 30, 40, 50, 60, 70, 80, 90, or 100% of the open reading frames in the genome of an organism.

In another aspect, the invention features a computer including a database of the invention and a user interface (i) capable of displaying one or more ligands that bind or modulate the activity of a target molecule whose record is stored in the computer or (ii) capable one or more target molecules that bind or have an activity that is modulated by a ligand whose record is stored in the computer. Exemplary databases include at least 10 records of target molecules, such as target molecules that have not been previously validated or target molecules of unknown biological function.

In another aspect, the invention provides an electronic database including at least 10^2 , 10^3 , 5×10^3 , 10^4 , 10^5 , 10^6 , 10^7 , 10^8 , or 10^9 records of compounds correlated to records of a phenotype in one or more biological assays that are effected by the compounds. The biological assay involves a cell or *in vitro*

WO 02/058533

PCT/US01/43348

sample that does not contain an exogenous copy of a nucleic acid encoding a protein that binds the compound or does not contain an exogenous reporter gene.

In another aspect, the invention features computer including the database of the above aspect and a user interface (i) capable of displaying one or more phenotypes in one or more biological assays for a compound whose record is stored in the computer or (ii) capable of displaying one or more compounds that effects a phenotype whose record is stored in the computer.

In another aspect, the invention provides electronic database including at least 10 records of target molecules correlated to records of an expression profile or activity of the target molecules. In another aspect, the invention features an electronic database including a plurality of records of target molecules that have not been previously validated as drug targets and/or target molecules of unknown function correlated to records of an expression profile or activity of the target molecules. In various embodiments of either database, the database includes records for at least 0.5, 1, 5, 10, 20, 30, 40, 50, 60, 70, 80, 90, or 100% of the proteins in the proteome of an organism, or on at least 10^2 , 10^3 , 5×10^3 , 10^4 , 10^5 , 10^6 , 10^7 , 10^8 , or 10^9 target molecules. In other embodiments, the database includes records for at least 0.5, 1, 5, 10, 20, 30, 40, 50, 60, 70, 80, 90, or 100% of the proteins in the proteome of an organism (e.g., the human proteome). In yet other embodiments, the database includes records for at least one protein expressed by an open reading frame for at least 0.5, 1, 5, 10, 20, 30, 40, 50, 60, 70, 80, 90, or 100% of the open reading frames in the genome of an organism.

In yet another aspect, the invention provides a computer including a database of the invention and a user interface (i) capable of displaying one or more expression profiles or activities of a target molecule whose record is stored in the computer or (ii) capable of displaying one or more target molecules that have an expression profile or activity whose record is stored in the computer. In various embodiments, the database includes at least 10 records of target molecules, such as target molecules that have not been previously validated as drug targets or target molecules of unknown function.

WO 02/052533

PCT/TOS91/43348

Any of the databases or computers can be used in any of the following methods. Exemplary uses of these databases include clustering of chemical scaffolds and types of active sites/proteins, global indexing of binding properties such as binding uniqueness and overlap, determining the specificity of scaffold
5 for a target, determining the potential toxicity of a compound, selecting a compound to probe a particular biology or pathology, selecting a target molecule responsible for the action of a particular compound, selecting a therapy based on pharmacogenomics, and selecting scaffolds to serve as leads for optimization of a drug.

10 In one such aspect, the invention features a method of identifying a target molecule associated with a phenotype of interest. This method involves using an electronic database including a plurality of records of phenotypes in a biological assay correlated to records of the ligands and their ability to cause or contribute to the phenotypes. A selection of a phenotype of interest is received, and one or
15 more ligands which contribute to the phenotype of interest are identified. An electronic database including a plurality of records of ligands correlated to records of the target molecules that bind the ligands or have an activity that is modulated by the ligands is used to identify one or more target molecules that bind or are modulated by the ligand(s) which contribute to the phenotype of
20 interest, thereby identifying one or more target molecules associated with the phenotype of interest. In one embodiment, the phenotype of interest is associated with a disease state, and the target molecule is determined to promote or inhibit the disease state. In one embodiment, the method is computer implemented.

25 In yet another aspect, the invention features a method of identifying a phenotype that is associated with a target molecule of interest. This method involves providing an electronic database including a plurality of records of target molecules correlated to records of the ligands and their ability to bind or modulate the activity of the target molecules, and receiving a selection of a
30 target molecule of interest. One or more ligands which bind or modulate the activity of the target molecule of interest are identified. An electronic database including a plurality of records of ligands correlated to records of phenotypes in

WO 02/058533

PCT/US01/43348

a biological assay caused by the ligands is provided and used to identify one or more phenotypes in a biological assay caused by the ligand(s), thereby identifying one or more phenotypes associated with the target molecule of interest. In one embodiment, the method is computer implemented.

5 In yet another aspect, the invention features a method of identifying a ligand that binds or modulates the activity of a target molecule of interest. This method involves providing an electronic database including at least 10 records of target molecules correlated to records of the ligands and their ability to bind or modulate the activity of the target molecules, and receiving a selection of a
10 target molecule of interest. One or more ligands which bind or modulate the activity of the target molecule of interest are identified. In various embodiments, the method includes comparing the chemical structures of two or more ligands which bind or modulate the activity of the target molecule of interest, thereby identifying functional groups in the ligands which promote the binding or
15 modulation of the target molecule of interest. In other embodiments, the method also includes comparing the chemical structures of two or more ligands which bind or modulate the activity of the target molecule of interest, thereby determining the frequency of one or more functional groups or scaffolds in the collection of the ligands. In other embodiments, one or more compounds that
20 have one or more functional groups that are present in two or more of the ligands for use in drug discovery or development or lead optimization. In one embodiment, the method is computer implemented.

In yet another aspect, the invention features a method of identifying a target molecule that binds or has an activity that is modulated by a ligand of
25 interest. This method involves providing an electronic database including at least 10 records of ligands correlated to records of the target molecules that bind or have an activity that is modulated by the ligands, and receiving a selection of a ligand of interest. One or more target molecules that bind or have an activity that is modulated by the ligand of interest are identified. In various
30 embodiments, the method includes comparing the chemical structures of two or more target molecules which bind the ligand of interest, thereby identifying

WO 02/058533

PCT/US01/43348

functional groups or domains in the target molecules which promote or contribute to the binding of the ligand of interest.

In yet another aspect, the invention features a method for determining the selectivity of a ligand of interest. This method involves providing an electronic database including at least 10 records of target molecules correlated to records of the ligands and their ability to bind or modulate the activity of the target molecules, and receiving a selection of a ligand of interest. The number of target molecules in the database that bind or are modulated by the ligand is determined, thereby determining the selectivity of the ligand of interest. In various embodiments, the ligand increases an activity of a target molecule, wherein the activity is associated with a disease state, an adverse side-effect, or toxicity and the ligand is eliminated from drug discovery or development, lead optimization, or development of an agricultural or environmental agent. In other embodiments, the ligand decreases an activity of a target molecule, wherein the activity is associated with a disease state, an adverse side-effect, or toxicity and the ligand is selected for discovery or development, lead optimization, or development of an agricultural or environmental agent. In one embodiment, the method is computer implemented.

In yet another aspect, the invention provides a method for selecting a therapy for a subject for the treatment, stabilization, or prevention of a disease or disorder. This method involves providing an electronic database including at least 10 records of target molecules correlated to records of the therapeutics and their ability to bind or modulate the activity of the target molecules, and determining a target molecule in the subject that has a mutation associated with the disease or disorder. A therapeutic is selected from the database that binds or modulates the activity of the target molecule and thereby treats, stabilizes, or prevents the disease or disorder. In other embodiment, the subject or a group of subjects having the mutation is selected for a clinical trial for the therapy or is classified in a particular subgroup for the clinical trial. In particular embodiments, the target molecule is a protein or nucleic acid. In one embodiment, the method is computer implemented.

WO 02/058533

PCT/US01/43348

In yet another aspect, the invention features another method for selecting a therapy for a subject for the treatment, stabilization, or prevention of a disease or disorder. This method involves providing an electronic database including at least 10 records of target molecules correlated to records of the therapeutics and their ability to bind or modulate the activity of the target molecules, and determining a target molecule in the subject that has a mutation associated with the disease or disorder. A therapeutic is selected from the database that does not bind or modulate the activity of the target molecule. In one embodiment, the mutation decreases the affinity of the target molecule for one or more therapeutics in the database and thus may decrease the efficacy of the therapeutic in that subject compared to subjects without the mutation. According to this embodiment, a therapeutic that binds a molecule other than the target molecule is selected. In other embodiment, the subject or a group of subjects having the mutation is excluded from a clinical trial for a therapeutic having decreased affinity for the mutant form of the target molecule, or the subject or a group of subjects is classified in a particular subgroup for the clinical trial. In yet other embodiment, the subject or a group of subjects having the mutation is selected for a clinical trial for a therapeutic that binds a molecule other than the target molecule, or the subject or a group of subjects is classified in a particular subgroup for the clinical trial. In particular embodiments, the target molecule is a protein or nucleic acid. In one embodiment, the method is computer implemented.

The invention also features improved methods for using mass spectrometry to determine whether a compound of interest is present in a sample. These methods may be used to identify ligands for particular target molecules.

In one such aspect, the invention provides a method of determining whether a compound of interest is present in a sample. This method involves determining or providing (i) reference mass spectra for two or more compounds from a library of compounds and (ii) a test mass spectrum of a sample including one or more compounds from the library. Whether or not one or more of the peaks of a reference mass spectrum are included in the test mass spectrum is determined, thereby determining whether the compound that generated the

WO 02/058533

PCT/US01/43348

reference mass spectrum is present in the sample. In various embodiments, the reference mass spectra are sequentially or simultaneously analyzed until all of the peaks in the test mass spectrum have been assigned to a compound. In other embodiments, the determination of whether or not the peaks of a reference mass spectrum are included in the test mass spectrum includes a sequential determination of whether the peaks of one or more reference mass spectrum are included in the test mass spectrum. In yet other embodiments, the determination of whether or not the peaks of a reference mass spectrum are included in the test mass spectrum is repeated until either (i) all of the peaks in the reference mass spectrum are determined to be present in the test mass spectrum, thereby determining that the compound that generated the reference mass spectrum is present in the sample, or (ii) a peak in the reference mass spectrum is determined to be absent in the test mass spectrum, thereby determining that the compound that generated the reference mass spectrum is not present in the sample.

In yet another aspect, the invention provides another method of determining whether a compound of interest is present in a sample. This method involves determining or providing (i) reference mass spectra of two or more compounds from a library of compounds and (ii) a test mass spectrum of a sample including one or more compounds from the library. One or more peaks of the test mass spectrum are analyzed to determine whether they are included in a reference mass spectrum. For a reference mass spectrum containing a peak that is present in the test mass spectrum, one or more of the other peaks in the reference mass spectrum are analyzed to determine whether they are present in the test mass spectrum, thereby determining whether the compound that generated the reference mass spectrum is present in the sample. In particular embodiments, the determination of whether the peaks in a reference mass spectrum are present in the test mass spectrum includes a sequential or simultaneous determination of whether the peaks of one or more reference mass spectrum are included in the test mass spectrum. In other embodiments, the determination of whether a peak in a reference mass spectrum is present in the test mass spectrum is repeated until either (i) all of the peaks in the reference mass spectrum are determined to be present in the test mass spectrum, thereby

WO 02/058533

PCT/US01/43348

determining that the compound that generated the reference mass spectrum is present in the sample, or (ii) a peak in the reference mass spectrum is determined to be absent in the test mass spectrum, thereby determining that the compound that generated the reference mass spectrum is not present in the sample.

5 In various embodiments of either of the above methods of determining whether a compound of interest is present in a sample, the mass spectrum of each compound in the library is determined. In yet other embodiments, at least one of the peaks in the reference spectrum is an isotope peak, a fragment peak, or a parent peak. In particular embodiments, the method involves determining
10 whether all of the peaks in a reference spectrum are present in the test mass spectrum. In other embodiments, the reference mass spectrum are contained in a database including records of one or more properties of mass spectra correlated to records of compounds that generate the mass spectra. In particular embodiments, the database contains data on one or more properties selected
15 from the group consisting of the mass to charge ratio of an isotope peak, the mass to charge ratio of a fragment peak, the mass to charge ratio of a parent peak, the intensity of an isotope peak, the intensity of a fragment peak, and the intensity of a parent peak. In still other embodiments, one or more of the steps for determining whether a peak in a test mass spectrum is present in a reference
20 mass spectrum are computer implemented.

In invention also provides a computer-readable memory having stored thereon a program for determining whether a compound of interest is present in a sample. This computer-readable memory includes computer code that receives as input mass spectrometry data including the mass to charge ratio for one or
25 more peaks in a reference mass spectra (i.e., the mass spectrum of an individual compound from a library of compounds). This computer-readable memory also includes computer code that receives as input mass spectrometry data including the mass to charge ratio for one or more peaks in a test mass spectra (i.e., the mass spectrum of a sample including one or more compounds from the library).
30 The computer-readable memory also has computer code that determines whether the peaks of a reference mass spectrum are included in the test mass spectrum.

WO 02/058533

PCT/US01/43348

thereby determining whether the compound that generated the reference mass spectrum is present in the sample.

In a related aspect, the invention features a computer-readable memory having stored thereon a program for determining whether a compound of interest is present in a sample. The memory includes computer code that receives as input mass spectrometry data including the mass to charge ratio for one or more peaks in a reference mass spectra (i.e., the mass spectrum of an individual compound from a library of compounds), and computer code that receives as input mass spectrometry data including the mass to charge ratio for one or more peaks in a test mass spectra (i.e., the mass spectrum of a sample including one or more compounds from the library). The memory also includes computer code that determines whether one or more peaks of the test mass spectrum are included in a reference mass spectrum, and computer code that determines whether all of the peaks in a reference mass spectrum are present in the test mass spectrum, thereby determining whether the compound that generated the reference mass spectrum is present in the sample.

The invention also features methods for the automated production of expression vectors or the automated production and purification of proteins.

In one such aspect, the invention features a method of producing two or more vectors encoding proteins of interest. This method involves robotically contacting a first nucleic acid encoding a first protein of interest with a first backbone nucleic acid in a robotic device under conditions that allow the their reaction, thereby producing a first vector encoding the first protein, and robotically contacting a second nucleic acid encoding a second protein of interest with a second vector nucleic acid in the robotic device under conditions that allow their reaction, thereby producing a second vector encoding the second protein. In some embodiments, the method also includes robotically contacting the first vector with a first cell under conditions that allow the insertion of the first vector into the first cell, and robotically contacting the second vector with a second cell under conditions that allow the insertion of the second vector into the second cell. In various embodiments, at least 3, 4, 5, 8, 10, 15, 30, 60, 90, or more vectors are produced simultaneously. In other embodiments, the backbone

WO 02/058533

PCT/US01/43348

nucleic acids are linearized expression vectors, and an insert encoding a protein of interest is ligated to the expression vector under conditions that generate a circularized expression vector containing the insert. In other embodiments, the first and second vectors or cells are contained in different flasks or wells in the robotic device. In other embodiments, the first cell expresses the first protein, and the second cell expresses the second protein. In yet other embodiments, the first protein and the second protein are purified as described in the aspect below. In other embodiments, the first cell and/or the second cell are bacteria such as *E. coli*, insect cells such as *Drosophila* cells, or mammalian cells such as Cos, HEK293, or CHO cells. In other embodiments, the first vector and the second vector are transferred from the first cell and the second cell to cells of another cell type, such as insect or mammalian cells, for the production of the first protein and the second protein. In other embodiments, a roller bottle system, Stir tank system, capillary cell culture system, or bioreactor is used to grow the cells. The first vector and/or the second vector can be used to produce protein to be used in any of the methods of the invention (e.g., to identify ligands that bind the protein).

One protein production and/or purification method of the invention involves expressing a first protein in a first cell under conditions that result in the secretion of the first protein into a first medium in a robotic device and expressing a second protein in a second cell under conditions that result in the secretion of the second protein into a second medium in the robotic device. The robotic device transfers the first medium to a first chromatography column and transfers the second medium to a second chromatography column. In one embodiment, the first protein and the second protein are isolated, thereby purifying the first protein and the second protein. In various embodiments, at least 3, 4, 5, 8, 10, 15, 30, 60, 90, or more proteins are purified simultaneously. In other embodiments, the first and second cells are contained in different flasks or wells in the robotic device. In other embodiments, the first cell and/or the second cell are bacteria such as *E. coli*, insect cells such as *Drosophila* cells, or mammalian cells such as Cos, HEK293, or CHO cells. In other embodiments, the first cell and/or second cell are transiently transfected Cos, HEK293,

WO 02/058533

PCT/US01/43348

Drosophila cells or CHO cells or stably transfected Cos, HEK293, CHO, *E. coli*, or Drosophila cells. In yet other embodiments, the first protein and/or the second protein are glycosylated in mammalian or insect cells. In various
5 embodiments, the first protein or the second protein naturally contain a secretion
signal or are genetically modified to contain a secretion signal so that they are
secreted by the cells into the medium. The first protein and/or the second protein
can be used in any of the methods of the invention (e.g., to identify ligands that
bind the protein). In other embodiments, the robotic device can be used to
10 contact the first protein and/or the second protein with a library of candidate
ligands to select ligands that bind the protein(s) using any of the methods
described herein. In yet other embodiments, the first protein and/or the second
protein are used as members of a library of target molecules that are robotically
connected with a small molecule of interest to select the target molecules that
bind the small molecule of interest using any of the methods described herein.

15 In various embodiments of any of the aspects of the invention, the ligand
binds a target molecule covalently or non-covalently. In other embodiments, the
ligand directly binds the target molecule or binds another molecule in the same
pathway as the target molecule and thereby activates or inhibits the target
molecule. In other embodiments, the ligand has a molecular weight of less than
20 5000, 4000, 3000, 2000, 1000, 750, 500, or 250 daltons. In other embodiments,
the ligand has less than 5, 4, 3, or 2 hydrogen-bond donors or less than 10, 8, 6,
4, or 3 hydrogen-bond acceptors. In yet other embodiments, the ligand has a c
logP of less than 4.15. In still other embodiments, the ligand is not FK506. In
other embodiments, the selected candidate ligands bind the target molecule with
25 a K_d of less than 1 fM, between 1 fM and 1 nM, between 1 nM and 1 μ M, or less
than 1 μ M. In other embodiments, the selected candidate ligands are subjected
to analysis by IR, MS, NMR, UV, amino acid sequencing, nucleic acid
sequencing, or a combination thereof. In other embodiments, an isotope or
fragment peak is used to identify a candidate ligand that has the same mass as
30 another candidate ligand in the library.

In various other embodiments of any of the aspects of the invention,
candidate ligands and/or the target molecules are in solution phase. In other

WO 02/058533

PCT/US01/43348

embodiments, the ligand or the target molecule is immobilized on a solid surface such as a bead or chip. In other embodiments, the assay medium is fractionated by chromatography. In particular embodiments, the complex is isolated using size exclusion (e.g., using silica or polymer resin), multimodal, bimodal, or biphasic chromatography (e.g., chromatography based on more than a single characteristic such as size exclusion and reverse phase, size exclusion and anionic exchange, size exclusion and cation exchange, or chromatography using an internal surface reverse phase (ISRP), GFF, or GFFII resin). Exemplary resins include diol, sepharose, sepharose, and polymethyl methacrylate. Other desirable resins are stable above 5, 50, 500, 5000, or 7000 psi. In particular embodiments, columns containing resins with different separation characteristics are combined in series. In other embodiments, column chromatography is used to isolate the complex, and the complex elutes from the column in less than 60, 30, 20, 15, 10, 5, 3, 2, or 1 minute; the void volume is less than 20, 15, 10, 5, 4, 3, 2, or 1 mL; or the column diameter is less than 5, 4, 3, 2, or 1 mm. In other embodiments, HPLC, spin columns, capillary chromatography, or filtration are used to isolate the complex. In other embodiments, a decrease in the UV absorbance of an HPLC or other chromatography peak corresponding to unbound ligand is used to detect a decrease in the amount of unbound ligand (and thus an increase in the amount of bound ligand). In still other embodiments, the complex of a target molecule and bound candidate ligands is subjected to a chromatography step that separates the bound ligands from the target molecule. In yet other embodiments of any of the aspects of the invention, an immobilized target is contacted with candidate ligand(s), and the support is washed with medium lacking candidate ligands and treated in manner that releases any bound ligands from the target. In still other embodiments, following exposure of the target to the candidate ligand(s), the support is washed with medium lacking target molecules, and treated in a manner that dislodges the candidate ligand molecules and any bound target molecules from the support. In other aspects, one, multiple, or all the steps in the method are robotically automated or computer implemented.

WO 02/058533

PCT/US01/43348

In still other embodiments of any of the aspects of the invention, the function or activity of a selected target is characterized by a chemical assay, biochemical assay, enzymatic assay, biological assay, or a combination thereof. In particular embodiments, the target function is characterized by an apoptosis assay, proliferation assay, necrosis assay, angiogenesis assay, invasion assay, or a combination thereof. In other embodiments, the candidate target molecules are isolated from biochemical extracts, cells, tissues, organisms, or recombinant sources. In yet other embodiments, a selected target molecule is identified using NMR, IR, UV, MS (e.g., MALDI-TOF, MALDI, single quad, triple quad, or electrospray MS or MS-MS), amino acid sequencing, or nucleic acid sequencing. In other embodiments, the candidate target molecule is a full-length protein or a fragment from a protein that is less than full-length. Exemplary targets include enzymes and receptors such as GPCRs, kinases, ion channels, nuclear receptors, proteases, phosphatases, and methylases. Targets may include molecules or classes of molecules for which therapeutically active compounds have or have not been previously developed.

It is noted that all of the embodiments of the various aspects of the invention for candidate ligands apply to small molecules of interest.

Herein, by "target molecule that has not been previously validated as a drug target" is meant a target molecule whose modulation has not been previously experimentally determined to promote or inhibit a disease state in an animal model of the disease, as described in a publication or public presentation. For example, unvalidated target molecules include molecules for which the activation or inhibition of the molecules or the decrease or increase in the expression level of the molecules has not been experimentally shown to modulate a disease state in an animal model of the disease. In contrast, validated drug targets include molecules for which increasing or decreasing the amount or an activity of the molecules has been experimentally determined to promote or inhibit a disease state in an animal model. Examples of validated targets include targets whose overexpression or inactivation due to a knockout mutation or other

WO 01/058533

PCT/US91/43348

gene silencing methods (e.g., antisense inhibition of gene expression) has been experimentally demonstrated to promote or inhibit a disease state in an animal model.

By "target molecule of unknown biological function" is meant a target molecule for which an activity has not been previously experimentally demonstrated, as described in a publication or public presentation. In various embodiments, the target molecule of unknown function is a nucleic acid or protein having less than 60, 50, 40, 30, 20, or 10% sequence identity to nucleic acids or proteins for which an activity has been experimentally demonstrated. In other embodiments, the nucleic acid or protein has not previously been assigned a putative function. Sequence identity is typically measured using sequence analysis software with the default parameters specified therein (e.g., Sequence Analysis Software Package of the Genetics Computer Group, University of Wisconsin Biotechnology Center, 1710 University Avenue, Madison, WI 53705). This software program matches similar sequences by assigning degrees of homology to various substitutions, deletions, and other modifications.

By "target molecule of unknown secondary or tertiary structure" is meant a target molecule for which the secondary or tertiary structure has not been previously experimentally determined, as described in a publication or public presentation. In some embodiments, the secondary or tertiary structure has not previously been predicted or modeled based on the known structure of a homologous molecule. In other embodiments, the location or tertiary structure of a binding site or active site in the target molecule has not been previously experimentally determined.

By "scaffold" is meant a core chemical structure that is contained in two or more different molecules in a library of candidate compounds. In various embodiments, at least 5, 10, 10², 10³, 10⁴, 10⁵, 10⁶, or more molecules in the library contain the scaffold. In some embodiments, the library contains at least 2, 2, 5, 10, 10², 10³, 10⁴, 10⁵, or more different scaffolds.

By "library" is meant a collection of 2, 5, 10, 10², 10³, 10⁴, 10⁵, 10⁶, 10⁷, 10⁸, 10⁹, or more different molecules. In various embodiments, each member of a library has a different mass. In other embodiments, at least 2, 5, 10, 15, 20,

WO 01/058533

PCT/US01/43348

30, 40, 50, or more of the members have the same mass or a mass that differs by less than 1, 0.5, 0.1, 0.05, or 0.01 daltons from the mass of another library member.

By "proteome" is meant all the proteins expressed by an organism. The proteome includes all of the alternative splice variants of a protein that are expressed by the organism.

By "purified" is meant separated from other components that naturally accompany it. Typically, a compound is substantially pure when it is at least 50%, by weight, free from proteins, antibodies, and naturally-occurring organic molecules with which it is naturally associated. In other embodiments, the compound is at least 75%, 90%, or 99%, by weight, pure. A substantially pure compound may be obtained by chemical synthesis, separation of the compound from natural sources, or production of the compound in a recombinant host cell that does not naturally produce the compound. Proteins and organic compounds may be purified by one skilled in the art using standard techniques such as those described by Ausubel *et al.* (Current Protocols in Molecular Biology, John Wiley & Sons, New York, 2000). The degree of purification compared to the starting material can be measured using standard methods such as polyacrylamide gel electrophoresis, column chromatography, optical density, HPLC analysis, or western analysis (Ausubel *et al.*, *supra*). Exemplary methods of purification include immunoprecipitation, column chromatography such as immunoaffinity chromatography, magnetic bead immunoaffinity purification, and panning with a phage-bound antibody.

The methods of the present invention have numerous advantages. For example, the methods allow the expression and purification of every protein in the proteome of an organism (e.g., the human proteome) and the identification of high-affinity, drug-like scaffolds for each protein. The methods also allow a theoretically unlimited number of candidate compounds and candidate scaffolds to be screened. Because the methods of the invention are so rapid and can be performed on such a large scale, they are useful for assaying target molecules that have not been previously validated as drug targets or target molecules of unknown biological function to select ligands that bind and/or modulate the

WO 02/058533

PCT/US01/43348

activity of the target molecules. In contrast, current methods for selecting ligands that bind a target molecule have been limited to target molecules that have been validated as drug targets. Thus, the present methods greatly expand the number of target molecules that can be assayed. Target molecules for which high affinity binders are selected can then be validated as drug targets.

Additionally, the methods of the invention allow candidate ligands that have the same mass to be distinguished. For example, mass spectral isotope and fragment peaks typically differ between ligands of the same mass. Thus, these peaks can be used to identify a candidate ligand even if it has the same parent peak as another candidate ligand in a library of compounds. This advantage allows the use of libraries containing multiple compounds of the same or similar masses.

The solution phase embodiments of the invention allow fluid phase binding to occur as it would in a serum or cell. In contrast to many current assays which measure a specific activity of the target protein, the methods of the present invention may be readily applied to any target in the proteome without customization. The methods also use a very small amount of reagents (such as <300 ug of each target for 200,000 compounds, and <35 ng of each compound for each target). The methods also allow a library of compounds to be screened without tagging or purifying individual members of the library before screening, thereby greatly decreasing the amount of time necessary to screen the library. The length of time required to screen libraries can also be reduced by using the automated embodiments of the present invention which allow multiple libraries and/or multiple targets to be analyzed in parallel.

Other advantages and embodiments of the invention will be apparent from the following detailed description and from the claims.

4. DESCRIPTION OF THE FIGURES

Figure 1 is an overview of the "genotype to phenotype" approach.

Figure 2 is an overview of the "phenotype to genotype" approach.

Figure 3 is a set of spectra illustrating the ability of P38 MAP kinase to isolate and extract a specific ligand with micromolar affinity.

WO 02/058533

PCT/US91/43348

Figure 4 is a set of UV spectra illustrating a P38 MAP kinase concentration dependant reduction of the 86002 peak but negligible reduction of the quinine peak in the HPLC separation of protein-bound compounds from free compounds.

Figure 5 is a set of mass spectra illustrating that the compound extracted from the mixture and released from p38 MAP kinase was identified as 86002.

Figure 6 is a list of the compounds in the 10 compound mixture and their molecular weights.

Figure 7 is a set of spectra demonstrating a P38 concentration dependent reduction of the 86002 peak but negligible reduction of the Colchicine peak or peaks representing the other compounds in the mixture during the HPLC separation of protein-bound compounds from free compounds. When the protein fraction was collected and the mass spectrum was determined, the spectrum included the peaks characteristic of 86002 at a level far higher than other peaks.

Figure 8 is a set of spectra illustrating a tubulin concentration dependent reduction of the Colchicine peak but negligible reduction of the 86002 peak or peaks representing the other compounds in the mixture during the HPLC separation of protein-bound compounds from free compounds. When the protein fraction was collected and the mass spectrum determined, the spectrum included the peaks characteristic of colchicine at a level far higher than other peaks.

Figure 9 is a list of the compounds in the 100 compound mixture and their molecular weights.

Figure 10 is a set of spectra illustrating that P38 MAP kinase binds and extracts a ligand with micromolar affinity (86002) from a 100 compound mixture in a specific and concentration dependent manner.

Figure 11 is a set of spectra illustrating that tubulin binds and extracts a hit (Colchicine) from a 100 compound mixture in a specific and concentration dependent manner.

Figure 12 is a set of UV spectra illustrating that excellent separation of the protein target from the unbound compounds in the 100 compound mixture is also achieved at higher flow rates.

WO 02/058533

PCT/US01/43348

Figure 13 is a set of spectra illustrating the ability of spin columns to separate a compound bound to a protein target from unbound compounds. This method was used to identify Colchicine as the predominant compound from the 100 compound mixture that bound tubulin.

5 Figure 14 is a schematic illustration of the steps in one embodiment of the Chemical Array Assay.

Figure 15 is a schematic illustration of an exemplary computer.

Figure 16 is an exemplary flow chart for one embodiment of the invention for identifying a compound in a sample.

10 Figure 17 is an graph illustrating the pairing of chemical scaffolds with protein targets which can be used to produce a chemical fingerprint of the human proteome.

Figure 18 is a schematic illustration of one embodiment for the automation and high throughput of methods of the invention to produce
15 ligand/target pairs.

Figure 19 is a schematic illustration of one embodiment for the high throughput production of ~2 milligrams of each of the ~90,000 proteins in the human proteome using automated cloning and production systems over a period of ~3 years at a rate of ~600 proteins per week.

20 5. DETAILED DESCRIPTION OF THE INVENTION

5.1. GENOTYPE TO PHENOTYPE

In one aspect, the present invention relates to methods of exposing protein or nucleic acid targets to a plurality of potential ligands, collecting
25 ligand—target pairs, and using the ligand(s) which bind the target to analyze the target's biological function. One embodiment is outlined in Figure 1. The method is used to determine the function of a target, which may be a target which has hitherto been unknown. Many other methods for selecting a candidate ligand that binds a target molecule are described herein. All of the embodiments
30 listed below in sections 5.1.1 to 5.1.5 can be used in any of the methods of the invention.

WO 02/059533

PCT/US01/43348

5.1.1. TARGETS

According to the present invention, a target molecule is the compound for which a binding or reacting molecule is sought. In preferred embodiments, the target is the species present at the highest concentration in the reaction vessel. In various preferred embodiments, the target is present at the same concentration as the ligand in the reaction vessel. In yet other preferred embodiments, the target is present at a higher or a lower concentration than the concentration of each ligand or the total concentration of the mixture of candidate ligands. In other preferred embodiments, the target is the species present at the lowest concentration in the reaction vessel. In one embodiment of the invention, the target is the species in the reaction vessel which has the highest molecular mass. A target may be a naturally occurring biomolecule synthesized *in vivo* or *in vitro*. A target may be comprised of amino acids, nucleic acids, sugars, lipids, natural products or combinations thereof. An advantage of the instant invention is that no prior knowledge of the identity or function of the target is necessary.

In a preferred embodiment of the invention, the target is comprised of amino acids, peptides, enzymes, proteins, antibodies or combinations thereof. In a first step, polynucleotides encoding the proteins of interest may be selected and introduced into an expression system. The polynucleotides may be selected by differential screening, subtractive hybridization, differential display, microarray expression analysis, representational difference analysis (RDA) or laser capture microdissection. The protein may be synthesized *in vivo* as in a bacterial plasmid, phage, transient cellular expression system or viral expression system. Alternatively, selected proteins may be synthesized *in vitro* by *in vitro* transcription and translation (e.g., Promega web site) or by common Fmoc oligopeptide synthesis chemistry. The expressed protein may be optionally purified and then exposed to a ligand library.

According to the invention, genes can be expressed from a complete cDNA or gene library of human or other species or a subset of genes selected for differential expression in a particular disease or upon a particular stimulus. Genes that are differentially expressed in diseased or stimulated cells and tissues

WO 02/058533

PCT/US91/43348

can be selected using but not limited to techniques such as subtractive hybridization, informatics, microarrays, SAGE, or laser capture microdissection. If partial sequences such as ESTs are recovered, full length tissue specific cDNAs may then be cloned from full length human cDNA libraries some of which are available from CLONTECH, STRATAGENE, Life Technologies, and NCBI. Between 20% and 60% of the genes being cloned in this way, depending upon the tissue, have not previously been identified and the functions of virtually every gene cloned have not been elucidated. In a preferred embodiment, these genes have been discovered by genomics. To produce proteins, the full length cDNAs may be tagged with hexahistidine (6his) inserted at the carboxyl terminal end and glutathione synthetase (GST) at the amino terminal end of the gene each with a protease cleavage site. Alternatively, the intein-based self cleaving tag by New England Biolabs may be used to avoid the need for protease treatment. These genes may be expressed and secreted into the supernatant by baculovirus, for example, using the Invitrogen® Schneider 2 *Drosophila* system with its his tag and bip protein leader, transfection using CaPO₄, and selection by hygromycin induced expression with copper sulfate, which can produce 5-10 mg/L of protein in the supernatant which can be purified over a nickel column. Non-limiting examples of alternative expression systems include Fast Bac or another baculoviral system or mammalian expression systems (CHO, COS, 293, etc.). *E. coli* may also be used for protein production but does not glycosylate proteins and the baculovirus system is as reliable and does glycosylate proteins. The resulting proteins can then be purified by Ni²⁺-NTA chromatography as a first purification step and glutathione affinity chromatography as a second step followed by specific protease removal by cleavage of the tags. If the intein based affinity system is used, no protease is required. The proteins can be expressed and purified using alternative techniques as well or the complete or partial protein may be expressed in phage or bound to a surface.

In another embodiment of the invention targets are comprised of RNA or DNA as oligonucleotides or polymucleotides. In one non-limiting embodiment of the present invention, nucleic acids to be introduced into an expression system

WO 02/058533

PCT/US01/43348

are identified by large scale sequencing of EST's. Oligonucleotide targets may be synthesized directly. Polynucleotide targets may be synthesized directly or prepared by amplification of a template polynucleotide, e.g., by PCR. The oligonucleotide or polynucleotide target may be optionally purified and then exposed to a ligand library.

In another embodiment of the invention, targets are comprised of simple or complex carbohydrates. In another embodiment of the invention, targets are comprised of lipids. In another embodiment of the invention, the target comprises natural products.

In another embodiment of the invention, the target may be derivatized. Non-limiting examples include biotin, fluorescein, digoxigenin, green fluorescent protein, radioisotope, his tag, magnetic bead, glutathione S transferase, photoactivatable crosslinkers or combinations thereof.

Target preparations may contain minor quantities of other compounds as a result of partial or incomplete purification of the desired component.

5.1.2. LIGANDS

According to the present invention, a ligand is any molecule which has the potential to bind to a target and/or exert an effect in a bioassay. In various embodiments of the genotype to phenotype approach, the ligand or the mixture of candidate ligands is present in the reaction vessel at a lower concentration than the target. In other embodiments of the phenotype to genotype approach, the ligand or the mixture of candidate ligands is present in the reaction vessel at the same concentration as the target. In still other embodiments of the genotype to phenotype approach, the ligand or the mixture of candidate ligands is present in the reaction vessel at a higher concentration than the target. A ligand may be comprised of amino acids, nucleic acids, sugars, lipids, natural products, natural product-like compounds or combinations thereof. A ligand may be created by any combinatorial chemical method. Alternatively, a ligand may be a naturally occurring biomolecule synthesized *in vivo* or *in vitro*. The ligand may be optionally derivatized with another compound. One advantage of this modification is that the derivatizing compound may be used to facilitate ligand-

WO 02/058533

PCT/US01/43348

target complex collection or ligand collection, e.g., after separation of ligand and target. Non-limiting examples of derivatizing groups include biotin, fluorescein, digoxigenin, green fluorescent protein, isotopes, polyhistidine, magnetic beads, glutathione S transferase, photoactivatable crosslinkers or combinations thereof.

5 Ligands should have low affinity for each other at the conditions under which the target is exposed to the ligand library.

 Ligand libraries are mixtures of ligands which differ from each other in mass, composition, structure or combinations thereof. The present invention contemplates such libraries which comprise at least 10 different ligands or at
10 least 100 different ligands or at least 1000 different ligands.

 The ligand library used to bind to the proteins can be derived from many sources. The invention includes the use of chemicals, proteins, peptides, antibodies, sugars, lipids, natural products, natural product-like compounds or any combination thereof. These may be prepared by organic synthesis,
15 combinatorial chemistry, recombinant DNA, biochemical extraction, purification, etc. In a preferred embodiment of the invention, natural product-like synthetic libraries are generated using diversity oriented chemistry (e.g., asymmetric split pool synthesis on beads or in solution, synthesized in parallel or in series), either combinatorial or medicinal chemistry. The subunits used in the
20 synthesis are preferably drug-like and are as highly diversified as possible. The units may be structurally rigid or flexible. The units may undergo chemical reactions that modify their own structures (e.g., rearrangement). The units may have functional groups added.

 Drug-like compounds may be made using different scaffolds with
25 different chemistries (e.g., organic, inorganic, peptide, protein, alkaloid, carbohydrate, lipids, natural product-like compounds). Drug-like compounds may incorporate spectral identifiers. Non-limiting examples of spectral identifiers include elements which resolve into characteristic isotope
 fragmentation patterns in mass spectroscopy (e.g., Cl, Br, N, H). Drug-like
30 compounds may also be made with compounds with unique fragmentation

WO 02/058533

PCT/US91/43348

patterns upon mass spectroscopy analysis (penicillin). The libraries can also be designed to facilitate other analytical and deconvolution techniques (e.g., IR FTIR).

In another embodiment of the invention, non-limiting examples of other libraries which may be used include commercially available libraries (e.g., PharmacoPeia, ArQule, and Chembridge), focused chemical libraries, peptides, peptides or proteins including the TAT, VP22 or ANTENNAPEDIA transduction signals, structurally flexible small molecules, natural products, sugars, and monoclonal antibodies. The subunits used in the synthesis are preferably drug like and are as highly diversified as possible.

Libraries of the invention may be tagged to facilitate ligand deconvolution and resynthesis after binding has been observed. Alternatively, the ligands can be deconvoluted without tagging. The ligands can be tested individually or in a mixture. Diverse libraries synthesized as a mixture in solution phase or on solid phase supports can be used. In one embodiment, the transduction peptides or variants thereof from TAT, VP22 or ANTENNAPEDIA can be crosslinked to a small molecule to enhance its ability to cross a membrane or barrier. Alternatively, a small molecule homologue of these peptides can be developed and linked to the same.

5.1.3. BINDING

According to the present invention, a ligand-target pair describes an affinity relationship between a ligand and target wherein the dissociation constant (K_d) is less than about 20 μ M, and preferably less than about 1 μ M.

The invention further contemplates ligand-target interactions where $K_d \leq 100$ nM or $K_d \leq 100$ pM or $K_d \leq 100$ fM. The interaction between the ligand and target may be covalent or non-covalent. The ligand of a ligand-target pair may or may not display affinity for other targets. The target of a ligand-target pair may or may not display affinity for other ligands.

According to the invention a reaction vessel is any container or surface in or upon which a target may be exposed to at least one of ligand. In a preferred embodiment of the invention, reaction vessels are arranged to facilitate high

WO 02/058533

PCT/US91/43348

throughput screening. This may be accomplished by using 96 or 384 well microtitre plates. Another possibility is depositing different target proteins on a glass slide at high density as illustrated by MacBeath *et al.*, 2000, Science 289:1760. In other embodiments of the invention the reaction vessel may be a column, resin, membrane, matrix, bead or chip.

The conditions under which the target is exposed to the ligand library may vary. Non-limiting examples include binding reactions where the temperature is less than about 5° C or from about 5° C to about 25° C or from about 25° C to about 40° C or over about 40° C. Further non-limiting examples include binding reaction conditions where the pH is less than about 5 or from about 5 to about 9 or over about 9. Further non-limiting examples include binding reactions in solutions which are comprised of water, an alcohol, an organic solvent or combinations thereof. Further non-limiting examples include binding reaction conditions where the additives may include ions, salts, detergents, reductants, oxidants or combinations thereof. A further non-limiting example includes binding reaction conditions where the target is immobilized. A further non-limiting example includes binding reaction conditions where ligands are immobilized. A further non-limiting example includes binding reaction conditions where targets are immobilized. A further non-limiting example includes binding reaction conditions where the target and the ligands are in solution.

A further non-limiting example includes binding reaction conditions where the ligand comprises a marker such as biotin, fluorescein, digoxigenin, green fluorescent protein, radioisotope, his tag, a magnetic bead, an enzyme or combinations thereof.

In one embodiment of the invention, the targets may be screened in a mechanism based assay. The mechanism based assay includes but is not limited to an assay to detect ligands which bind to the target. This may include a solid phase or fluid phase binding event with either the ligand, the protein or an indicator of either being detected. Alternatively, the gene encoding the protein with previously undefined function can be transfected with a reporter system (including but not limited to β -galactosidase, luciferase, green fluorescent

WO 02/058533

PCT/US01/43348

protein, etc.) into a cell and screened against the library ideally by a high throughput or ultra high throughput (e.g., 1560 well per plate of chip) screening or with individual members of the library. In an alternative embodiment of the invention other mechanism based binding assays may be used. These include
 5 other assays including biochemical assays measuring an effect on enzymatic activity, cell based assays in which the target and a reporter system (e.g., luciferase or β -galactosidase) have been introduced into a cell, and binding assays which detect changes in free energy. Binding assays can be performed with the target fixed to a well, bead or chip or captured by an immobilized
 10 antibody or resolved by capillary electrophoresis. The bound ligands may be detected usually using colorimetric or fluorescence or surface plasmon resonance. In the column based binding assay, the binding may be performed in a well or other vessel, on a gel, etc.

While there are a number of ways these assays can be done, following
 15 inductive thought, only the chemicals which bind to the protein target are relevant and can teach its function. In addition, the fluid phase more accurately reflects the true biological conformation. Furthermore, in the reaction both the protein and the chemicals preferably are not tagged, decreasing the problem that the protein has been constrained in some way by coupling to a plate or a bead or
 20 the ligand is not in the same fluid phase conformation which it will be in the cell or the blood. Consequently, in a preferred embodiment of the invention, 1 to 20,000 ligands (with 1000 to 10,000 preferred) may be mixed together with 1 ng to 1 mg of each protein (with 0.1 to 100 μ g preferred) in a small volume (1 μ L to 1 mL, with preferred range of 0.1 μ L to 100 μ L) to have a 0.1 μ M to 100 μ M
 25 concentration with a preferred range of 0.1 μ M to 10 μ M. In particular embodiments of the invention, by looking at only the 1 to 500 ligands which would be expected to bind to each protein with micromolar to nanomolar affinity, one avoids having to screen millions of combinations individually. This overcomes the need to tag the library in any other way than the molecules own
 30 mass, isotope pattern or fragmentation pattern, because mass spectroscopy can

WO 02/058533

PCT/US01/43348

resolve and identify the possible 1 to 5 hits per well. Alternatively, IR and/or FTIR can be used alone or in combination with mass spectroscopy to resolve and identify hits.

5 5.1.4. LIGAND-TARGET SEPARATION AND LIGAND IDENTIFICATION

In a preferred embodiment of the invention, ligand-target pairs are separated from unbound ligands and unbound targets by liquid chromatography, ligand-target pairs are separated from each other in a second liquid
10 chromatography step, and ligands which bind are identified by mass spectroscopy. In various embodiments of the invention, the solution phase binding may occur in a well, tube or column. Capillary electrophoresis, and/or other detection methods may be used to deconvolute ligands from the library. Particularly, HPLC and mass spectroscopy or capillary electrophoresis and mass
15 spectroscopy can measure the molecules with extreme sensitivity. In addition, this technique can be done in extremely small volumes which is critical to optimally utilize the small amounts of each member of the chemical library. For example, less than 20,000 ligands from the chemical library may be pooled with the protein for binding again in each well in 96 well plates at $\leq 10 \mu\text{M}$ in
20 approximately 100 μL and 1 μg of protein. In a preferred embodiment, HPLC is performed in 96 well plates with cartridges to serve as the columns for each well. In another embodiment, the separation is performed in parallel in 384 well, 1536 well, or 10,000 or greater well formats using column, wells, cartridges, chips, or filters. Alternatively, this may be performed in a standard HPLC column, spin
25 column, or other column. The first cartridge/column may be a gel permeation or size exclusion or gel filtration (e.g., S25 like resin, Pharmacia) to hold the unbound molecules in the resin but allow the bound ligand and protein to pass through. A small sample volume is desired (preferably 1 to 100 μL or less) yet this procedure may dilute the sample by one or more orders of magnitude. It is
30 helpful, therefore, to use a small and narrow column (preferably having a diameter of 1 to 2 mm or less and a length of 5 to 200 mm (Rocket Column, Biorad or Pharmacia columns) to minimize dilution of the sample. Capillary

WO 02/058533

PCT/US01/43348

Liquid Chromatography can also be used. This resin separates the protein along with small molecules bound to it with high affinity ($K_d \leq 1.0 \mu M$). The next cartridge/column would use a hydrophobic or hydrophilic reverse phase HPLC resin, the choice of which depends upon the hydrophobicity of the ligand library being used: C18 (silica hydrophobic- used with less hydrophobic ligand) C8 column (more hydrophilic, used for more hydrophobic ligands), a cyanocolumn (use for more hydrophilic ligands) or SB8U from Agilent which can be used for either hydrophilic or hydrophobic ligands. These reverse phase HPLC methods separate the bound small molecule ligands from the protein and concentrate the small molecules and protein sample via resin binding. Subsequently, the small molecules may be eluted from the protein and the resin and the eluents may be collected in a 96 well plate. Providing one knows the amount of the starting material, affinity may also be measured in this step. Alternatively, competition studies can be done at a later time to quantitate binding affinity.

These eluents may then be transferred to a mass spectrometer and characterized. This may be done robotically in real time potentially even in the 96 well format perhaps using either a parallel multiple channel microchip system or a parallel spray interface. Alternatively, chip based MALDI TOF Mass spectrometry may be used. In this case, the protein fraction from the column (spia, HPLC, capillary, other) can be spotted onto a chip or a filter in a 96 well or greater format. The Omnisflex or Autoflex MALDI instruments from Bruker Daltonics automatically describe and analyze each of the samples from 160 sample and 1536 sample formats, respectively. Nonlimiting forms of mass spectrometry that may be used include electrospray, ion trap, Fourier Transform, MALDI, single or triple quadrupole in single MS, MS-MS, or MS-MS-MS formats.

Eluents may be characterized using a software package for use with the mass spectrometer supplemented with information about the ligand library used. Mass spectroscopy may be used to identify compounds by direct detection of its mass. However, mass spectroscopy may also be used to detect compounds, scaffolds or linkers containing elements which resolve into characteristic isotope patterns (e.g., ^{35}Cl , ^{15}N , 2H) or compounds having unique fragmentation patterns

WO 02/058533

PCT/US01/43348

(e.g., penicillin). For example, chlorine-containing compounds will be comprised of ^{35}Cl and ^{37}Cl which will produce two mass peaks, 2 AMU apart with a 3:1 intensity ratio. Similarly, bromine-containing compounds will be comprised of ^{79}Br and ^{81}Br which will produce two mass peaks, 2 AMU apart with a 1:1 intensity ratio. This approaches may be used as an alternative to or in combination with true molecular weight to identify a compound.

Mass spectroscopy enables the mass, isotope, and fragmentation pattern to be determined so accurately that, coupled with software, the exact member of the library may be identified except for the isomer. Following this the theoretically expected 500 or so micromolar to nanomolar hits can be pulled from the original library and synthesized in a target scale. If the molecule is a peptide, it can be fused to the TAT transducing sequence which allows proteins to cross the cell membrane.

In another embodiment of the invention, ligands are characterized by IR or FTIR in addition to or instead of mass spectroscopy analysis. These techniques permit identification of ligand functional groups or substitutions (e.g., hydroxyl or amino groups). Used in combination with mass spectroscopy, this may facilitate differentiation between ligands of identical molecular weight.

According to the invention, the dissociation constant (K_d) of the ligand-target pair should be less than about 100 μM and preferably less than about 10 μM . While not dispositive, the dissociation constant (K_d) of the ligand-target pair is one factor which may guide those skilled in the art in determining the utility of a ligand in determining target function and as a drug lead. Thus, the invention contemplates but does not necessarily prefer ligand-target pair interactions where the dissociation constant (K_d) is less than about 1 μM or less than about 100 nM or less than about 10 nM or less than about 1 nM or less than about 100 pM or less than about 10 pM.

If no hits or a low number of hits with reasonable affinity are found, a structural or chemical gap in the structural diversity of the chemical library may have been identified. In such a case, target directed synthesis can be employed to fill in that gap. If low affinity binders are found, the binding can be repeated with a library containing photoactivatable (or other) linkers on one of the

WO 02/658533

PCT/US01/43348

functional domains. After the first column when only the protein and molecules binding to it are present, the photoactivation step can be performed, after which the small molecules can be eluted by reverse phase HPLC. In this way, the target has been used as a template and because two molecules which bound with a low affinity linked together will have an increased affinity for the target. In a preferred embodiment, the increase in affinity is 2 to 100 fold.

5.1.4.1. Exemplary Chemical Array Assay Experimental Methods and Results

10 Methods for HPLC Based Assay

Drug-like chemical compounds representing a collection of drug-like chemical scaffolds (Sigma-Aldrich, ICN, Calbiochem) were weighed and mixed to a final concentration of 20 μ M each in 50 mM ammonium acetate pH 7, 10% methanol. 1 μ M to 20 μ M tubulin or P38 MAP kinase (Sigma) were dispensed into HPLC low volume sample cuvettes (Waters) and mixed with 0.5 μ M to 20 μ M compounds. After mixing and a 15 minute 37°C incubation, the cuvettes were placed on ice and injected into the HPLC (Waters 2690) using an autosampler (Waters) onto a 150mm X 2.1mm ID Pinkerton GFF II column (Regis Technologies) for dual size exclusion and phase separation with a 50 mM ammonium acetate, 10% methanol running buffer. The protein target and bound compounds eluted in the column void volume as detected using a Diode array detector and most of the compounds absorbed well at a 243 nm frequency. In some cases, using low concentrations of each compound (0.5 to 5 mM) and fewer than 10 compounds which could be easily separated from one another, it was possible to titrate in the two protein targets and observe a corresponding titration in the level of UV absorbance of the specific compound known to bind one of the protein targets but not to nonspecific control compounds.

We optimized the column dimensions and the choice of resin to maximize the separation of the compounds bound to the protein targets from the unbound compounds. Resins which elute protein in the void volume and small column diameters and lengths which minimize the void volume were used. Such columns minimize the amount of dilution of the protein sample and minimize the

WO 02/058533

PCT/US01/43348

time required for each assay, thereby minimizing the amount of bound compound that dissociates from the protein (as governed by the $K_{d,2}$ rate). These features enabled the use of minimal amounts of reagents, as well as sensitive detection methods. The column lengths were such that the protein eluted in less than 2 to 3 minutes. A number of HPLC columns, including the Regis 150 mm x 2.1 mm GFF II column, a 1.0 mm x 100 mm YMC Diol column, a 2.1 mm x 150 mm Phenomenex Polyhydroxymethacrylate (Polysep) column, and a Jirdi 2.1 x 150 mm Divinyl Benzene column, were tested. Similarly, other running buffers were tested in which the salt and methanol concentration were varied, and the ratio of protein target to small compounds in the binding reaction was varied from 1000:1 to 1:1000. Resins representative of different classes were tested for their ability to separate the protein fraction from the drug-like small molecule compounds, and to minimize the cycle time for all of the compounds to elute from the column. These characteristics of the columns are determined by surface properties and limitations on flow rates due to resins collapsing under backpressure. Being silica based and thus resistant to pressure, the YMC diol column had a cycle time of under 10 minutes but was only able to separate approximately 50% of the compounds in the 100 compound mixture listed in Fig. 9 from the protein. The Phenomenex Polyhydroxymethacrylate column was able to separate approximately 80% of the compounds in the 100 compound mixture from the protein, and required a methanol gradient to achieve elution of many of the small molecule compounds; it tolerated a relatively low flow rate (0.18 ml/min) because of the inability to tolerate backpressures over 600PSI. The cycle time for the Phenomenex column was 1.5 hours with the gradient, and 35 minute for a subset of compounds (15% of the total) which could be isolated without the gradient. Other polymer based columns [e.g., polyhydroxymethacrylate (Phenomenex, Shodex, Waters), polymethylmethacrylate (Shodex, TosohBiosep), Sepharose/Sephadex/Superoose (Amersham Pharmacia Biotech)] also only tolerated relatively low flow rates. The Jirdi DVB columns are divinyl benzene polymer columns, which were operated at high pressure (4000PSI) and undesirably bound the protein as well as the compounds, thus giving no separation in the buffer system used. Other

WO 02/058533

PCT/US01/43348

buffer systems are expected to allow separation of the protein from the unbound compounds. Different columns and resins were also combined in series, increasing the percentage of compounds separated from the protein but also increasing the cycle time. In applications where a longer cycle time (e.g., over
5 10 minutes per run) is acceptable, any of the above columns or a series of the above columns may be used.

For shorter cycle times, other columns may be used. For example, the Regis GFF II column separated the protein fraction from 97% of the compounds tested. Its pressure rating of 8000PSI was above that of the HPLC (Waters
10 2690) used in these assays, which was operated at a pressure of 6000PSI. The cycle time of this resin was demonstrated to be easily less than 8 minutes and could be further decreased by using a faster flow rate in an HPLC that tolerates pressures up to 8000PSI. The GFF II resin and GFF resin are internal surface reversed phase resins which were developed by Thomas Pinkerton for the direct
15 analysis of drugs and drug metabolites in serum without interference by protein adsorption. The resins consist of a porous silica support with a hydrophilic external surface and hydrophobic internal pores accessible only to molecules with a molecular weight less than 12,000 daltons. These surfaces are produced by bonding the tripeptide glycine-phenylalanine-phenylalanine (GFF) or
20 glycidoxylpropyl-glycine-phenylalanine-phenylalanine (GFF II) to the silica surfaces. The GFF or GFF II bonded beads are then treated with the exopeptidase, carboxypeptidase A, which has a molecular weight (35,000 daltons) large enough to exclude it from the pores resulting in the cleavage of the
phenylalanine-phenylalanine portion from the outer surface. This treatment
25 allows the glycine or glycidoxylpropyl to be exposed intact on the outer surface making the outer surface hydrophilic but leaving the original tripeptide intact on the inner surface, thereby making the inner surface hydrophobic (as described, for example, by the manufacturer's packaging insert). The catalogue number of the column with the GFF II resin that was used is 288-4. Other columns with
30 other catalogue numbers that are packed with these resins are also available from Regis technologies and can also be used. The outer surface thus prevents large molecules from entering the inner layer through size exclusion and hydrophilic

WO 02/058533

PCT/US01/43348

interactions. Small molecules enter the inner surface which is comprised of the hydrophobic support which retains and separates the compounds based upon hydrophobic interactions. Given the short cycle times and the degree of separation that can be achieved with the GFF II resin, the GFF II column was used for subsequent assays; however, other resins can also be used.

Protein fractions from the HPLC columns were dissociated with 1%TFA, and a 100nL sample was injected onto a reverse phase column (Waters Symmetry Shield) to separate the compounds that had been bound to the protein. The compounds were eluted using an acetonitrile gradient past a UV detector and into a TOF mass spectrometer (Micromass LCT). The background signal was subtracted from each sample using controls containing the protein in the absence of compounds, and the mass spectrum was determined at cone voltages high enough to achieve fragmentation of the compounds (20 to 80 volts). In other mass spectrometry instruments, fragmentation can be achieved in a collision cell. The fragmentation pattern which is characteristic for each compound consists of the larger parent peak and other peaks representing fragments of the chemical compound or their isotopes. The fragmentation pattern of the compound(s) released from the protein target was compared to the characteristic fragmentation pattern observed for a compound standard to identify the compound(s) that bound the protein target. Alternatively, one or more characteristic isotope(s) of the parent peak representing the molecular weight of the compound was compared with the standard to identify the compound that bound the protein target. In another alternative analysis, the parent peak representing the molecular weight of the compound was itself compared with the standard to identify the compound. Sometimes, the combination of these methods was also used to identify the compound. Similar methods were applied under MS conditions which did not induce fragmentation of the compound, resulting in a mass spectrum containing peaks representing the molecular weight of the compound (e.g. the parent peak) and its isotopes.

30

WO 02/058533

PCT/US01/43348

Results from HPLC Based Method

SKBR6002 is a ligand with micromolar affinity for the P38 MAP kinase protein target. P38 MAP kinase (5 μ M) was mixed with 5 μ M 86002 and separated by HPLC on the Diol column (Fig. 3). The protein fraction was collected and analyzed by mass spectrometry. The parent peak, fragments, and isotopic peaks in the spectrum corresponded to the 86002 standard indicating that the P38 MAP kinase isolates and extracts a specific ligand with micromolar affinity.

SKBR6002 and quinine monohydrochloride (a nonspecific control compound) were mixed together to a final concentration of 5 μ M each (Fig. 4). Increasing amounts of P38 MAP kinase protein (final concentrations 0, 2.5, 5 and 10 μ M) were mixed with the compound mixture at a final concentration of 5 μ M each, and the protein was separated by HPLC on the Diol column. The UV spectrum demonstrated a P38 concentration dependent reduction of the 86002 peak but negligible reduction of the quinine peak.

When the P38 protein fraction was collected at the mid-point in the titration (5 μ M P38 MAP kinase + 5 μ M mixture of Quinine and 86002) illustrated in Fig. 4, the compound extracted from the mixture and released from the protein was identified as 86002, and not quinine, based on the parent peak, fragments, and isotopic peaks in the mass spectrum of the released compound (Fig. 5).

A mixture of equal amounts of 10 drug-like compounds including 86002 and colchicine was prepared (Fig. 6). Increasing amounts of P38 MAP kinase protein (final concentrations 0, 3.5, and 5 μ M) were mixed with the 10 compound mixture at a final concentration of 0.5 μ M of each compound, and the protein was separated by HPLC on the GFF II column (Fig. 7). The UV spectrum demonstrated a P38 concentration dependent reduction of the 86002 peak but negligible reduction of the Colchicine peak or peaks representing the other compounds in the mixture. When the protein fraction was collected and the mass spectrum was determined, the spectrum included the parent and isotopic peaks characteristic of 86002 at a level far higher than other peaks.

WO 02/058533

PCT/US01/43348

Increasing amounts of tubulin protein (final concentrations 0, 5, and 20 μM) were mixed with the 10 compound mixture at a final concentration of 0.5 μM of each compound, and the protein was separated by HPLC on the GFF II column (Fig. 8). The UV spectrum demonstrated a tubulin concentration dependent reduction of the Colchicine peak but negligible reduction of the 86002 peak or peaks representing the other compounds in the mixture. When the protein fraction was collected and the mass spectrum determined, the spectrum included the peaks characteristic of Colchicine at a level far higher than other peaks.

A mixture of equal amounts of 100 drug like compounds including 86002 and Colchicine was prepared (Fig. 9). P38 (2 μM) was mixed with the 100 compound mixture at a final concentration of 20 μM of each compound, and the protein was separated from the unbound compounds using the GFF II HPLC column (Fig. 10). The protein fraction was collected, the compound were released from the protein and mass spectrum was determined. The spectrum contained a peak characteristic of 86002 at a level far higher than other peaks. Thus, P38 MAP kinase binds and extracts a ligand with micromolar affinity (86002) from a 100 compound mixture in a specific and concentration dependent manner. The mass spectrum background appears to be comparable to that generated using only 10 compounds (Fig. 7), indicating that the assay should be scalable to larger numbers of compounds (e.g., 1000's to 10,000's of compounds). For example, these methods may be used to analyze a library of over 10, 20, 40, 50, 75, 100, 200, 500, 1000, 2000, 5000, 10000, or more compounds or more chemical scaffolds.

Tubulin (5 μM) was mixed with the 100 compound mixture at a final concentration of 5 μM of each compound, and the protein was separated from the unbound compounds using the GFF II HPLC column (Fig. 11). The protein fraction was collected, the compound were released from the protein, and the mass spectrum was determined. The spectrum showed the peaks characteristic of colchicine at a level far higher than other peaks. Thus, tubulin binds and extracts a hit (Colchicine) from a 100 compound mixture in a specific and concentration dependent manner. The mass spectrum background appears to be

WO 02/058533

PCT/US01/43348

comparable to that generated using the 10 compound mixture (Fig. 8), indicating that the assay should be scalable to larger numbers of compounds (e.g., 1000's to 10,000's of compounds). For example, these methods may be used to analyze a library of over 10, 20, 40, 50, 75, 100, 200, 500, 1000, 2000, 5000, 10000, or more compounds or more chemical scaffolds.

One way to increase the speed of the assay is to increase the flow rate (Fig. 12). The limiting factor affecting the maximum flow rate a column can withstand is generally the backpressure which the resin can tolerate before it collapses. One of the reasons the GFF II resin was selected is its ability to sustain pressures up to 8000PSI compared with most size exclusion gels (e.g., Sepharose, Superose, Superdex, polymethylmethacrylate, polyhydroxymethacrylate, etc.) which have maximum back pressures of 100-1500PSI. At high flow rates, the GFF II column still achieved excellent separation of the protein from the 100 compound mix.

Spin Column Chromatography Methods

Drug-like chemical compounds representing a collection of drug-like chemical scaffolds (Sigma-Aldrich, ICN, Calbiochem) were weighed and mixed to a final concentration of 20 uM each in 50mM ammonium acetate pH 7, 10% methanol. 5 uM to 20 uM bovine serum albumin (BSA) or tubulin (Sigma) were dispensed into HPLC low volume sample cuvettes (Waters) and mixed with 5 uM to 20 uM compounds. After mixing and a 15 minute 37°C incubation, the cuvettes were placed on ice. 50 uL of the 100 compound mixture listed in Fig. 9 was then layered on top of a MicroSpin G-25 (Amersham Pharmacia Biotech) spin column which had been previously equilibrated with two washes of binding buffer (i.e., each wash involved adding 200 uL of 50 mM ammonium acetate, 10% methanol buffer, and spinning the buffer through the column into a 1.5 ml. microfuge tube (Eppendorf) at maximum setting in a microfuge (Eppendorf) for 30 seconds to a minute). Such spin columns are generally used to desalt and exchange buffer for DNA probes after labeling, though G-25 is one of the classic size exclusion resins with a 25KD molecular weight cut off. The spin column was then placed in a 1.5 ml. microfuge tube (Eppendorf) and spun for 30 seconds

WO 02/058533

PCT/US01/43348

at maximum setting in the microfuge (Eppendorf). Alternatively, a vacuum can be used to pull solution through the spin column which is particularly useful when spin column/cartridges are arrayed in the 96 well format and a vacuum manifold is used to pull the solution through the column into a 96 well plate.

5 In the case of BSA, the 50 μ L solution in the bottom of the microfuge tube was loaded onto the HPLC, the UV spectrum was visualized and compared with an equivalent amount of the BSA/100 compound mixture before separation. In the case of tubulin, 25 μ L of the solution at the bottom of the microfuge tube was dissociated with 1%TFA and injected onto a reverse phase column (Waters Symmetry Shield), and the compounds were eluted using an acetonitrile gradient
10 past a UV detector into a TOF MS (Micromass LCT). Background was electronically subtracted from each sample using controls containing the protein in the absence of compounds and the mass spectrum was determined at cone voltages high enough to achieve fragmentation of the compounds (20 to 80
15 volts). In other mass spectrometers, such fragmentation can be achieved in a collision cell. The fragmentation pattern which is characteristic for each compound consists of the larger parent peak and other peaks representing fragments of the chemical compound or their isotopes. The fragmentation pattern of the compound(s) released from the protein target was compared to the
20 characteristic fragmentation pattern observed for a compound standard to identify the compound(s) that bound the protein target. Alternatively, a characteristic isotope of the parent peak representing the molecular weight of the compound was compared with the standard to identify the compound that bound the protein target. In another alternative analysis, the parent peak representing
25 the molecular weight of the compound was itself compared with the standard to identify the compound. Sometimes, the combination of these methods was also used to identify the compound. Similar methods were applied under MS
conditions which did not induce fragmentation of the compound, resulting in a
30 mass spectrum containing peaks representing the molecular weight of the compound (e.g., the parent peak) and its isotopes.

WO 02/058533

PCT/US91/43348

Results from Spin-Column Chromatography Based Methods

5 μ M Bovine serum albumin (BSA, Sigma) was mixed with the 100 compound mixture at a final concentration of 5 μ M of each compound (Fig. 13).

- 5 Half (50 μ L) of the mixture was layered on top of a Micro-Spin G-25 column and centrifuged. The protein containing fraction was collected at the bottom of the microfuge tube. When the initial protein/compound mixture was compared with the protein/compound mixture after separation using the spin column separation method, a significant purification of the protein was observed based on UV absorbance. When the same protocol was applied to a mixture of 20 μ M tubulin and 20 μ M of the 100 compound mixture and the mass spectrum was determined for the eluted protein-containing fraction, the spectrum showed the peaks characteristic of Colchicine at a level far higher than other peaks. Although the background peak was slightly higher than that observed using the HPLC column separation (Fig. 14), the speed and scalability of this spin column separation make it highly attractive. For example, these methods may be used to analyze a library of over 10, 20, 40, 50, 75, 100, 200, 500, 1000, 2000, 5000, 10000, or more compounds or more chemical scaffolds.

20 5.1.4.2. Exemplary Methods for the Use of Pattern Recognition Software to Identify Isolated Ligand(s)

- The present invention provides methods for using pattern recognition analysis of a mass spectrum to identify a compound from a mixture that has been isolated using a protein target and any of the separation techniques described herein.

- 25 In these methods, mass spectrometry fragmentation patterns are determined for many or all of the compound present in the initial mixture of candidate compounds. Alternatively, isotope or other mass spectrometry patterns are determined for these compounds (e.g., M+1 or M+2 isotope peaks). The mass spectrometer sorts the compounds, their isotopes, and/or their fragments on the basis of their mass to charge ratio, denoted m/z. The mass spectrometry conditions can be adjusted so that most or all of the peaks represent

WO 02/058533

PCT/US01/43348

molecules having a charge of +1 (or -1), so that the value of some of the peaks is equal to the mass of the parent compound, an isotope, or a fragment of the parent compound (*i.e.*, $m/z = m/1 = m$). In some cases, other mass spectrometry conditions can be used so that some or all of the peaks represent molecules
 5 having a charge of +2 or greater (or -2 or lower), so that the value of some of the peaks is less than the mass of the parent compound, an isotope, or a fragment because the mass to charge ratio is less than the mass of the molecule (*e.g.*, $m/z = m/2$). Thus, the mass spectrometry patterns consist of mass spectral peaks corresponding to masses (or mass to charge ratios if the charge on the molecules
 10 is greater than one) of the parent compounds, their fragments, and/or their isotopes.

The mass (or mass to charge ratio) of each of these peaks is entered into the database of an information retrieval system. The mass spectrum of a compound of interest that was released from a protein target is generated, and
 15 then pattern recognition software is used to compare this pattern with those contained in the database. A match positively identifies the compound of interest. In one embodiment, peaks corresponding to two, three, or more of the most characteristic masses (compound 1: peaks A, B, and C; compound 2: peaks D, and E; *etc.*) are entered into the database for each of the compounds in the
 20 initial mixture. Software (*e.g.*, MassLynx, version 3.5 from Micromass) is used to search the mass spectrum of the compound(s) released from a protein target for peak A followed sequentially by a search for peaks B, C, D, F, *etc.* The presence of a particular peak is entered into a second database to indicate that the peak is present in the mass spectrum. In another possible method, the searches
 25 for particular peaks in the mass spectrum are performed in any order. Iterative search commands may also be used to analyze the mass spectrum. For example, if peak A corresponding to a particular compound is present in the mass spectrum, then the mass spectrum can be analyzed to determine whether another peak (*e.g.*, peak B) characteristic of the same compound is also present in the
 30 mass spectrum. Alternatively, if a peak characteristic of a particular compound is not present in the mass spectrum, then the mass spectrum can be analyzed to determine whether a peak (*e.g.*, peak D) characteristic of another compound is

WO 02/058333

PCT/US01/43348

present in the mass spectrum. In yet another alternative method, multiple peaks are searched together by overlaying a macro program over MassLynx. The peaks identified as present are compared with those in the first database from the compounds in the initial mixture to identify the compound(s) released from the protein target. Fig. 16 A contains an exemplary flow chart illustrating the steps for some embodiments of these methods.

In another embodiment, two, three, or more masses (or mass to charge ratios) corresponding to the most characteristic peaks of the mass spectrometry pattern are entered into the database for each compound in the initial mixture. In an exemplary method, this database uses a Microsoft Excel or Oracle program. Once the mass spectrum for the sample released from the protein target is determined and the two or three main peaks in the mass spectrum (e.g., the two or three peaks with the highest signal) are located, a search is performed on the database for the initial compound mixture using the masses (or mass to charge ratios) corresponding to those peaks. For example, the values of the masses can be used in the "Find" command of these programs to search for candidate compounds that produce peaks of that mass. The combination of masses identified in the search thus identifies the compound(s) present in the sample.

In a yet another embodiment, the intensity of the signal at a particular mass (or mass to charge ratios) is used to positively identify a compound. This technique is particularly applicable if the pattern being used is an isotype pattern. In this case, a database of compounds in the mixture is generated that contains both the mass as well as the intensity of each of the two or three most characteristic peaks. This information is then collected for the sample of interest. The search function of the database program is used to search for the correlated mass and intensity parameters. A match positively identifies a compound present in the sample.

In various embodiments for any of the methods of the present invention for the identification of one or more compounds of interest (e.g., compounds released from a target), one or more mass spectral peaks corresponding to one or more fragments of a compound and/or one or more mass spectral peaks corresponding to one or more isotopes of a compound is used to identify the

WO 02/058533

PCT/US01/43348

compound. In other embodiments, the parent peak is used in the identification of the compound. In various embodiments, the parent peak is the only spectral peak used in the identification of a compound. In yet other embodiments, the parent peak is used in conjunction with one or more peaks corresponding to a fragment or an isotope in the identification of a compound. In still other
5 embodiments, a parent peak is not used in the identification of the compound. In other embodiments, the compound is a component recovered from a mixture of at least 5, 10, 20, 40, 50, 75, 100, 200, 500, 1000, 2000, 5000, 10000 or more compounds that were contacted with a target of interest. In other embodiments, the compound is a component recovered from a mixture of compounds that
10 includes at least 5, 10, 20, 40, 50, 75, 100, 200, 500, 1000, 2000, 5000, 10000 or more different chemical scaffolds. In particular embodiments, a parent peak is used in the identification of a compound from a mixture of compounds that includes at least 5, 10, 20, 40, 50, 75, 100, 200, 500, 1000, 2000, 5000, 10000 or
15 more different chemical scaffolds.

Any of the methods described herein may be implemented using virtually any computer. Fig. 15 shows such an exemplary computer system. Computer system 2 includes internal and external components. The internal components include a processor 4 coupled to a memory 6. The external components include a mass-storage device 8, e.g., a hard disk drive, user input devices 10, e.g., a
20 keyboard and a mouse, a display 12, e.g., a monitor, and usually, a network link 14 capable of connecting the computer system to other computers to allow sharing of data and processing tasks. Programs are loaded into the memory 6 of this system 2 during operation. These programs include an operating system 16, e.g., Microsoft Windows, which manages the computer system, software 18 that encodes common languages and functions to assist programs that implement the methods of this invention, and software 20 that encodes the methods of the
25 invention in a procedural language or symbolic package. Languages that can be used to program the methods include, without limitation, Visual C++ from Microsoft. In preferred applications, the methods of the invention are programmed in mathematical software packages that allow symbolic entry of equations and high-level specification of processing, including algorithms used
30

WO 02/058533

PCT/US01/43348

in the execution of the programs, thereby freeing a user of the need to program procedurally individual equations or algorithms. An exemplary mathematical software package useful for this purpose is Matlab from Mathworks (Natick, MA). Using the Matlab software, one can also apply the Parallel Virtual Machine (PVM) module and Message Passing Interface (MPI), which supports processing on multiple processors. This implementation of PVM and MPI with the methods herein is accomplished using methods known in the art. Alternatively, the software or a portion thereof is encoded in dedicated circuitry by methods known in the art.

10

5.1.5. ANALYSIS OF TARGET FUNCTION

To systematically classify target function, the hits for each target may be screened in cell and tissue based assays representing each of the major molecular mechanisms in disease pathogenesis. Where the target is originally selected based on differential expression analysis, assays which are particularly relevant to that differential expression are preferred (e.g., a proliferation assay would be particularly relevant where the target arose from differential expression analysis of carcinoma cells). This panel of assays includes but is not limited to assays to detect and/or measure: apoptosis, proliferation, ischemia/necrosis, inflammation, fibrosis, angiogenesis, metabolic signaling, infection and development/differentiation. By focusing on pathogenic pathways and studying disease specific and cell specific targets, novel targets for a number of therapeutic areas may be identified. The goal of this panel is to screen for small molecule/protein members of the molecular pathways leading to significant diseases including but not limited to chronic degenerative diseases (e.g., Alzheimer's disease, osteoarthritis, osteoporosis), metabolic diseases (e.g., diabetes, obesity), inflammatory diseases, cancer, cardiovascular (e.g., coronary artery disease, hypertension, congestive heart failure, cardiomyopathy, chronic renal failure) and infectious (e.g., viral, bacterial, protozoan, and mechanisms of drug resistance). The assays are designed such that the same assay can be used in cells first with follow up in tissue biopsied from patients with the disease. To identify potentially toxic molecules, necrosis assays may be performed on all

15

20

25

30

WO 02/058533

PCT/US01/43348

molecules. The standard industry microtitre plates of 96 wells provide sufficient scale to conduct these phenotypic screens though high throughput and ultra high throughput formats are not precluded. Assays may be performed on cell lines, primary cell culture, tissue biopsies, tissue models, *in vivo* animal models, or other organisms. In a preferred embodiment, the bioassays are performed using human cell lines and tissues. According to other embodiments, the bioassays may be performed using cells, tissues, organs or whole organisms of any species. Though ligands can be pooled in these assays, it is useful that each phenotypic assay be performed with one species of molecule per well to avoid agonist and antagonist interactions which may mask the phenotypic effect. The assays include but are not limited to allowing the diseased cell or tissue to enrich for genes which may be relevant to disease or a therapeutic response.

Although applications of the invention toward target identification in cancer, diabetes and stimulation of cells with TGF β are described in the examples, the approach set forth above can be broadly applied to any disease, cell stimulus, biological modulator or condition. Other assays than those described and those for other molecular pathways relevant to diseases can also be used. By taking this approach starting with genes up-regulated or down-regulated in diseased cells relative to normal cells or tissues or in cells in the presence of an agonist or antagonist (or partial of each) one is enriching for targets with specificity and a good therapeutic index. By crossing this specificity with molecular mechanisms in disease pathogenesis, one is enriching for targets which may be therapeutic. By sequentially combining a biochemical binding assay which selects hits in a highly efficient manner from large libraries and using these hits in a low throughput high quality phenotypic bioassay reflective of the human disease, one can determine the function of the gene.

5.2. PHENOTYPE TO GENOTYPE

In an alternative series of embodiments, the present invention relates to a method of screening a plurality of potential ligands in at least one bioassay, selecting ligands which produce a change in phenotype in a bioassay, and using the ligand to screen candidate targets to identify the particular target(s)

WO 02/058533

PCT/US01/43348

responsible for the altered phenotype. In various preferred embodiments, individual species of ligands are separately screened in bioassay(s). A ligand which produces a change in phenotype in a bioassay may be exposed to a plurality of potential targets under conditions which permit ligand-target
 5 interaction. In various preferred embodiments of the invention, the target is a peptide or protein and each peptide or protein target is associated with a polynucleotide which encodes that target (e.g., by phage display or cell surface display). Selected targets and their corresponding polynucleotides are collected. The DNA sequence encoding targets which are proteins may be sequenced,
 10 cloned, and validated. The differential expression of these targets may then be studied in human disease tissue biopsies particularly where the molecular mechanism of the phenotype may be phenotypically relevant. Similarly the ligand may be studied in diseased tissues and/or *in vitro* or *in vivo* models of these diseases. One embodiment is outlined in Figure 2. As noted above, the
 15 embodiments listed in sections 5.1.1 to 5.1.5 can be used in any of these methods.

High throughput phenotype cell based assays according to the invention differ from high throughput screening methods as they are currently practiced. The typical high throughput screen is a mechanism based assay where the gene
 20 for a validated target is transfected into a cell line with a reporter system (e.g., green fluorescent protein, luciferase, etc.) and members of a chemical library are screened for activation of the reporter. Instead of conducting this type of screen, the present invention focuses on looking for a significant change in phenotype in cell lines without predetermining the molecular target in a bioassay. These
 25 bioassays are designed to look for ligands which modulate an important biological stimulus or an important pathogenic mechanism. Non-limiting examples include apoptosis, proliferation, ischemia, necrosis, inflammation, fibrosis, invasion, angiogenesis, metabolism, infection and embryogenesis. In addition, individual pathways of cellular stimuli with pleiotropic effects can be
 30 blocked by antisense, translocating peptides, antibodies or other techniques to identify targets which are more specific in their effect. In this way we achieve an association of ligands from the library (as described above) with a phenotype

WO 02/058533

PCT/US01/43348

in a bioassay. Assays for molecular mechanisms in disease including but not limited to those described above may be adapted to high throughput screening.

Although applications of the invention toward target identification in cancer are discussed herein, the invention can be broadly applied to any disease, cell stimulus or condition. Other assays than those described related to biological stimuli and those for other molecular pathways relevant to diseases or biology can also be used. By sequentially combining a bioassay in which a ligand is associated with a particular phenotypic change of interest and using these hits to select for the target in a protein or peptide display library, one can clone the gene for and identify the target. The differential expression of the target in human disease tissue may then be studied. In addition, the specificity of a ligand's effect in an *in vitro* or *in vivo* bioassay may reveal the utility of that ligand in modulating a biological affect or treating a particular disease.

15 5.3. MAPPING MOLECULAR SIGNALING PATHWAYS

Once a number of genes have been shown to be involved in a particular molecular pathway of disease pathogenesis the targets can be mapped within the molecular pathway relative to one another and to known members of the pathway. The ligands binding to the different proteins may be derivatized with photoactivatable crosslinkers and used to position each member in the pathway. For example, one member of a pathway is first labeled (e.g., GFP). Next, members of the pathway are exposed to ligands derivatized with functional groups which may be crosslinked. Then, the mixture is exposed to the crosslinking stimulus. Lastly, the selected member of the pathway is collected using the label (e.g., GFP) and any compounds which have become associated with it are identified. This may be repeated stepwise to identify earlier or later pathway members. These methods have the advantage of not requiring the prior identification of the binding sites for the ligands or the determination of the secondary or tertiary structure of the target molecule prior to crosslinking.

30 Pathway members may then be used as targets in ligand screens. By comparing the phenotype of each ligand which selectively binds each pathway member, positional information about each pathway member relative to others

WO 02/058533

PCT/US01/43348

may be obtained. This information can be used to validate and select the best target for a given disease indication and eventually select the best therapy through pharmacogenetic based diagnosis.

5 5.4. OPTIMIZATION OF LEADS

The present invention provides a method for optimizing leads and increasing the hit ratio. The term "lead" as used herein refers to a ligand with pharmaceutically desirable properties. Preferably the molecule would be considered a "small" molecule in the art, for example having a molecular weight between 50 Da and 5000 Da. The method has broad application, but is particularly useful for obtaining ligands which interfere with protein-protein interactions.

Because a large number of chemical leads may be characterized at the biochemical and phenotypic levels, a structure activity relationship may be established to serve as a basis for lead optimization. If molecules with similar activities are identified, the structure activity relationship (SAR) can be determined. A target directed synthesis technology can be employed to crosslink molecules binding close to each other indicating if their activity is mediated through the same active subsite on the protein or through different subsites on the protein target. In one embodiment, one of the molecules contains a photoactivatable crosslinker, or one molecule contains a reactive group that is reactive with a group on a second molecule. In this way additional different functional subsites on the target can be mapped and different mechanisms can be interpreted from the phenotypic findings with molecules binding to those subsites (e.g., agonist vs. antagonist). Photoactivatable crosslinkers on one of the functional groups of the ligand scaffold may be used to link ligands bound to the target thus using the target molecule as a template.

In this process, small molecule A and small molecule B can be mixed alone or in the presence of other nonbonding small molecules with the target (s) and a bifunctional crosslinker capable of reacting with both A and B in which one functional group is protected and the other is free. Alternatively, A can be reacted with a crosslinker, and the resulting product can be reacted with B.

WO 02058533

PCT/US99/43348

Functional groups can include any reactive group, including, but not limited to, amines, carboxylic acid, nitrile, and halides. The same or different functional groups can be on A or B. In one example of a pair of small molecules A and B that can react with each other, A contains an amine functional group, and B contains a crosslinker with a carboxylic acid, an activated ester, and anhydride, an acylhalide, or any other group which can react with the amine in an acylation or an alkylation reaction. Linkers can include a molecule which only contains two functional groups or contains a component in between the functional groups including, but not limited to, polyethylene glycol. Exemplary protective groups include amine protecting groups such as BOC, Fmoc, or benzyl. The CBZ protecting group can be used to protect carboxylic acids benzylester, allylester, and nitriles. In one embodiment, protective groups are photoactivated to deprotect a functional group, such as Nitrobenzyl or azo groups. In another embodiment, linkers containing functional groups which do not react with proteins and compounds which do not contain the functional groups on proteins (i.e., amines, carboxylic acids, alcohol, and SH groups) are used. In an example, the compound contains or is modified to contain a halide (e.g., Cl). A linker containing double bonds, triple bonds, halides, or aromatic groups can then be linked to the compound through a Heck coupling reaction or a Suzuki reaction resulting in a linkage of the linker with the compound without reacting with the protein. Such chemical compounds are available from Aldrich. Linkers and protective groups for the above reactions are available from Advanced Chemtech and Novobiochem among others. This linking may increase the affinity of binding to the target in a preferred embodiment between 2 and 100 fold or more. Thus, a superior lead with higher affinity can be obtained. This approach can also be used to further enhance the structural diversity of a chemical library in a target directed and biologically relevant way.

WO 02/058533

PCT/US01/43348

6. GENOTYPE TO PHENOTYPE

6.1. EXAMPLE 1: BREAST CANCER6.1.1. TARGETS

A biopsy is first collected from at least one breast cancer patient. Laser capture microdissection and ANRNA or RT PCR may be used in conjunction with microarray analysis to isolate genes which are differentially expressed in the cancerous cells. For example, these techniques may be used to identify transcripts which are present in cancer cells at levels more than 2-fold higher than non-cancerous cells in the same biopsy. Alternatively, the genes may be overexpressed in non-cancerous cells. Genes may further be selected for those which are expressed at such levels in a significant fraction of patients tested.

Tissue may be embedded in Tissue Tek OCT medium (VWR), frozen in liquid nitrogen, and sectioned in a cryostat. Sections may be mounted on uncoated glass slides and stored at -80° C. Slides may be fixed in 70% ethanol for 30 s, stained with H&E followed by 5 s dehydration steps in 70%, 95%, and 100% and a 5 min dehydration step in xylene. After air drying, the sections may be laser microdissected using the PixCell i and II LCM system (Arcturus Engineering). 5 X 10⁴ each of morphologically normal breast epithelial cells, malignant invasive breast carcinoma cells and malignant metastatic breast carcinoma cells (e.g., from the axillary lymph node) may be captured. The total RNA may be isolated from each of these cell populations by transferring a transfer film with adherent cells into guanidinium isothiocyanate at room temperature, extracting with phenol/chloroform/isoamyl alcohol, and precipitating with sodium acetate and 10 µg/µL glycogen in isopropanol. The RNA pellet may then be resuspended and treated with 10 units DNase (GeneHunter) in the presence of RNase inhibitor (Life Technologies) for 2 hours at 37° C. Following extraction and precipitation, the pellet may be resuspended in 27 µL of RNase free water. ANRNA or RT PCR may be performed followed by sequencing. Sequences identified by this technique which are EST's may be used to select a full length cDNA from a cDNA library (CLONTECH). These cDNA's may be enriched in diseased but not normal cells/tissues but their function may be unknown.

WO 02/058533

PCT/US01/43348

Selected cDNA's may be each tagged with hexahistidine (6his) inserted at the carboxy terminal end and glutathione synthetase (GST) at the amino terminal end of the gene each with a protease cleavage site. These genes may be cloned into a *Drosophila* expression system vector with the bip protein leader, co-transfected with hygromycin vector into *Drosophila* using CaPO_4 . Cells may be maintained in selective media and gene expression may be induced with copper sulfate (Invitrogen). After 48 hours, supernatant containing 5-10 mg/L of each protein may be collected. The resulting proteins may then be purified from the supernatant by Ni^{2+} -NTA chromatography, as a first purification step, and glutathione affinity chromatography, as a second step, followed by specific protease removal by cleavage of the tags. Up to milligram quantities of each protein may be recovered.

6.1.2. BINDING LIGAND TARGET PAIR SELECTION AND

15 LIGAND IDENTIFICATION

Diverse chemical, natural product-like and peptide combinatorial libraries containing up to 2 million ligands may be synthesized in a pooled fashion in fluid phase. In addition, natural product libraries (Tetragen, Yonsei), and chemical libraries (Anqule, Coelocath) may be purchased. From 1,000 to 10,000 ligands may be mixed together with 1 μg of protein in a volume of up to 100 μL to have a 1 μM concentration in the well of a 96 well plate. After a 30 minute incubation on ice, the samples may be loaded into 96 well plates with cartridges to serve as IIPLC columns for each well (Waters 2790 IIPLC). The first cartridge/column may be a size exclusion resin (G25 Pharmacia) to hold the unbound molecules in the resin but allow the bound ligand and protein to pass through. A small and narrow column (e.g., 2 mm length x 5 mm diameter Rocket Column, Biorad) is used to minimize dilution at this step. The next cartridge/column used is a hydrophobic or hydrophilic reverse phase HPLC resin, the choice of which depends upon the hydrophobicity of the ligand library being used. For example, a hydrophobic C18 silica column may be used with less hydrophobic ligands, while a hydrophilic C8 column may be used for more hydrophilic ligands. Another example is the SB8U column from Agilent which

WO 02/058533

PCT/US01/43348

may be used for either hydrophilic or hydrophobic ligands. The reverse phase HPLC may concentrate the small molecules and protein by allowing them to bind onto the resin after which the small molecules may be eluted from the protein and the resin. The eluants containing the small molecules may be collected in a 96 well plate. These eluants may then be transferred to the mass spectrometer (Micromass Quattro LC) and the spectra determined using the Masslynx, MAXENT software (Micromass). In this way theoretically up to 100 ligands per protein may be deconvoluted such that the exact member of the library may be identified except for chirality. Specifically, mass spectroscopy can be used to detect isotopes of compounds or fragmentation patterns any of which can be used as an alternative or in combination with true molecular weight to identify a compound. In addition, IR or FTIR analysis may be performed to identify ligand functional groups or units. Each ligand may then be synthesized or on a larger scale. Peptide ligands may be fused with the TAT transducing sequence.

The affinity of the ligands identified will depend in part on the concentration of the library used in the screen, but should range from at least nanomolar to micromolar. The actual affinity of each ligand may be determined by competition studies. These ligands may then be tested in bioassays.

20

6.1.3. BIOASSAYS

Where the cDNAs are selected based on their differential expression in cancer cells, the ligands may be tested in assays which detect or measure apoptosis, proliferation, necrosis, angiogenesis, inflammation, or metastatic tumor invasion. According to the invention, assays are designed using models which are as close to the human disease as possible (e.g., pathological tissue biopsies, *in vitro* tissue models, *in vitro* disease models, human cell lines) and which are based upon cell lines and are easily applied to primary tissue from human pathology samples. These assays may be developed using tissue from mice transgenic for a gene known to be involved in cancer, *bc1-2*. Human breast cancer cell lines which may be assayed include: MCF-7, NCI/ADR HS578T, MDA-MR-22231/ATCC, MDA-MB-4335, MDA-N, BT-549, T-47D (NCI,

30

WO 02/058533

PCT/US91/43348

ATCC). Other cell lines and tissues may also be used. Non-limiting examples of bioassays are shown in Table 1.

Table 1: Bioassays in cell lines, human tissue biopsies, and human tissue biopsies transplanted into host (e.g., nude mouse).

Pathogenic Mechanism	Bioassay in breast, colon, lung, and prostate cell lines (e.g., breast cancer, MCF-7, NCI/ADR HS578T, MDA-MB-232/ATCC, MDA-MB-435, MDA-N, BT-20, T-47D)
Apoptosis	1.5 hour <i>in vitro</i> incubation with ligand then stain with FITC Annexin V; DAPI stain nuclear morphology confirmation.
Necrosis	8 hour incubation with ligand (in nude mouse); vital dye stain with propidium iodide or TOPO-3, confirm with MTT assay.
Proliferation	2 hour incubation with ligand then stain with FITC anti-PCNA; confirm with BRDU.
Angiogenesis	Incubate tumor in nude mouse with ligand, stain with fluorescein factor VIII related antigen to measure endothelial cell density; confirm in migration of cultured human dermal microvasculature endothelial cells towards β -PGF.
Inflammation	2 hour incubation with ligand and measure TNF, INF, IL-4, IL-2, IL-10, TGF β , VCAM, MCP-1 via ELISA.
Invasion	30 hour incubation of cells labeled with CFSE dye in Matrigel cell invasion chamber; confirm by study in nude mice.
Fibrosis	48 hour incubation with ligand followed by fibronectin ELISA assay or immunohistochemistry.
Metabolism	2 hour incubation with insulin and ligand then measure glucose levels; test in 3T3-L1 adipocytes and L6 myoblast cell lines followed by type II diabetes compared to normal patient fat biopsies.
Development/Differentiation	Incubate ligand with either MHC class II-negative cells or single pluripotent ML-IC cells and assess cell fate by cytological and immunological techniques according to either Inaba K <i>et al.</i> , 1993, PNAS 90:3038 or Purzel M <i>et al.</i> , 1999, Blood 93:3750.

WO 02/058533

PCT/US01/43348

6.1.3.1. APOPTOSIS

Apoptosis may be assayed using a cell membrane phosphatidyl serine binding dye (FITC Annexin V; alternative dyes such as Cy5.5 may also be used). Selected ligands for each of the proteins identified in the binding assay may be tested for an effect on apoptosis on various cell lines. From 2×10^3 to 2×10^6 cells may be plated in each well of a 96 well plate and medium containing 1 μ M to 10 μ M of each ligand is added to wells in triplicate. Minimally, a negative (no ligands) and a positive (*bcl-2* inactive ligand) control are also performed. After 1.5 hours, FITC Annexin is added to the wells, incubated with the cells for 15 minutes and, after 3 washing steps, the level of fluorescence is determined using a plate reader.

The assays may be demonstrated to be transferable from cells to tissues by using *bcl-2* expressing cells and tissues from *bcl-2* transgenic mice (Charles River). Ligands which induce apoptosis may be tested on fresh tumor biopsies from breast cancer patients. One advantage of using primary tissue biopsy is that the assay may be performed within two hours of tissue collection, i.e. before the tissue has begun showing the changes associated with ischemia. Small pieces of tumor biopsy may be plated in wells of a 96 well plate and the same assay as above is repeated with each sample in duplicate. After, the fluorescence is read, the samples may be stained with DAPI staining (Molecular Probes, Eugene Oregon) and nuclear morphology may be assessed under a fluorescence microscope for nuclear condensation and fragmentation for confirmation. Alternatively, the classic TUNEL (terminal deoxynucleotidyl transferase mediated biotinylated deoxyuridine triphosphate nick end labeling) method to label DNA strand breaks may be used.

6.1.3.2. PROLIFERATION

Cell proliferation may be assayed by exposing cells to a fluorescein labeled anti-PCNA antibody (e.g., PC-10, Santa Cruz Biotechnology) which binds to proliferating cell nuclear antigen (PCNA). Selected ligands for each of the proteins identified in the binding assay may be tested for an effect on proliferation on cell lines. From 2×10^3 to 2×10^6 cells may be plated in each well

WO 02/053533

PCT/US01/43348

of a 96 well plate. Medium containing 1 μ M to 10 μ M of each ligand may then be added to wells in triplicate. Minimally, a negative (no ligands) and a positive control are also performed. After 2 hours, FITC anti-PCNA may be added to the wells, incubated with the cells for 15 minutes and, after 3 washing steps, the level of fluorescence may be determined using a plate reader. The PCNA assay has already been used in cells and in tissues (Kulldorff M et. al., 2000, J. Clin Epidemiology 53:875). Ligands which inhibit proliferation may be tested on fresh tumor biopsies from breast cancer patients. Small pieces of tumor biopsy may be plated in wells of a 96 well plate and the same assay as above repeated with each sample in duplicate. After the fluorescence is read, the samples may be assessed under a fluorescence microscope to confirm that the cells whose proliferation indeed is being affected are the cancer cells.

In a second approach cell proliferation is classically measured looking at BRDU or 3 H- thymidine uptake. According to a third approach, cells may be labeled with the CFSE dye (5-and-6 carboxyfluorescein diacetate succinimidyl ester). As the cells proliferate over 7 to 8 generations, the dye is diluted. A fourth approach uses a fluorescence-based AtinPhos assay to measure endogenous enzyme acid phosphatase may be used to measure cell numbers. Other methods for detecting cells undergoing proliferation may be used, including 7-ADD (7-amino-actinomycin-D) which is used to determine the stage of proliferation or by staining with the Ki67 antibody.

6.1.3.3. NECROSIS

Techniques to detect necrosis include but are not limited to the classic techniques of DNA binding dyes such as propidium iodide or TOTO-3. Alternatively, a colorimetric methylthiazole tetrazolium (MTT) assay for the mitochondrial enzyme release can also be used to determine cell viability. In a preferred embodiment of the invention, cell viability is determined using the DNA binding dyes propidium iodide and TOTO-3. Conducting these assays in cell lines may enable one to distinguish between necrosis and apoptosis which will facilitate distinguishing ligands have specific effects from ligands which are

WO 02/058533

PCT/US01/43348

broadly cytotoxic. This distinction may also be facilitated by performing necrosis and apoptosis assays in parallel. Selected ligands for each of the targets identified in the binding assay may be tested for an effect on necrosis of the cell lines. From 2×10^5 to 2×10^6 cells may be plated in each well of a 96 well plate and medium containing 1 μ M to 10 μ M of each ligand is added to wells in triplicate. Minimally, a negative (no ligands) and a positive control are also performed. After 8 hours, propidium iodide or TOTO 3 is added to the wells, incubated with the cells for 15 minutes and after 3 washing steps, the level of fluorescence is determined using a fluorescent plate reader.

Necrosis may be a difficult assay to transfer to tissue biopsies because it is generally assayed after at least 8 hours and there is a lot of necrosis due to ischemia in tissue biopsies after such an interval providing a high background. To overcome this problem, human biopsy tissue may be transplanted into nude mice, thereby preventing ischemia induced necrosis during the 8 hour assay period. To insure that growth in the nude mouse does not alter the tumor, a tumor, grown in a nude mouse for 1 month, may be explanted and tested in the short term apoptosis and proliferation as outlined above. The tumor may also be viewed histologically and compared with the fresh tumor explant to assess differences. The ligands which bind to the same target and induce necrosis in 50% of the cases may be injected into the tumor in the animal, collected after 8 hours, and stained with propidium iodide. Histological examination may reveal that the tumor cells are undergoing necrosis while the other cells in the biopsy are not.

6.1.3.4. ANGIOGENESIS

The *in vitro* assay used to test for a pro or anti-angiogenic effect assays the migration of cultured human dermal microvascular endothelial cells towards β -FGF or bovine serum albumin (negative control) with increasing concentrations of angiostatin as an inhibitory control and increasing concentrations of the ligands in different wells (Clonetics, San Diego; Polverini PJ et al., 1991, Methods in Enzymology 198: 440). Angiogenesis is also a longer term event so modeling in human biopsies will absolutely require growth

WO 01/058533

PCT/US01/43348

in nude mice. Should ligands with an anti-angiogenic activity be discovered in the future, they may be assayed by daily injection into the tumor for 3 to 5 days and subsequent removal and staining with Fluorescent anti-Factor VIII related antigen to measure endothelial cell density.

- 5 Other models for angiogenesis are contemplated by the invention. *In vivo* models include implantation of hydropellets with the test molecules on them implanted into the avascular rat cornea (cornea micropocket assay). Growth of vessels from the limbus to towards the pellet at 7 days is scored as a positive response which can be negated by the removal of the angiogenic or anti-angiogenic protein by antibody on protein A beads (Poverini PJ et. al., 1991, Methods in Enzymology 198: 440). These vessels can be characterized as to the density, length and luminal sizes of the vessels. A similar assay can also be performed in the mouse eye (L. Smith, Children's Hospital, Boston). Angiogenic molecules can also be tested *in vivo* in the rabbit model of hindlimb ischemia (Shty KG et. al., 1998 Circulation 98:2081). Other *in vitro* tissue modeling systems include endothelial cells in 3 dimensional culture where they form tubular structures that resemble immature capillaries (Springhorn et. al., 1995, *In vitro Cell Dev Biol Anim* 31, 473; Sierra-Honigsmann MR et. al., 1996, Science 281:1683). Smooth muscle cell recruitment can be measured using anti-smooth muscle actin immunohistochemistry.
- 10
- 15
- 20

6.1.3.5. INVASION

- Tumor invasion may be assayed using the a basement membrane cell invasion chamber which is a chamber coated with Matrigel extracellular matrix.
- 25 The matrix coats the wells used to separate one chamber from the other in 24 well plates (Becton Dickinson Labware). Selected ligands for each of the proteins identified in the binding assay may be tested for an effect on invasion on the cell lines. Cells labeled with CFSE dye can be measured by FACS or used to follow cell fate *in vivo*. Alternatively, cells may be labeled with ³H-thymidine or another marker. About 2x10⁵ labeled cells may be plated in each well and medium containing 1 μM or 10 μM of each ligand is added to the top half of the wells in triplicate. After 30 hours in a CO₂ incubator, the membrane
- 30

WO 01/05353

PCT/US01/43348

chambers may be rinsed 3 time on both sides with DMEM/0.1% BSA and the top surface is scrubbed with a cotton swab. The amount of dye present in the bottom well may be determined using a fluorescent plate reader. In positive wells, the membrane can be cut out and the number of cells on the bottom can be counted. Ligands affecting tumor invasion in this *in vitro* assay may be further tested *in vivo* by histological analysis of human tumor biopsies in nude mice.

6.1.3.6. DEVELOPMENT AND/OR DIFFERENTIATION

Various assays to test the effect of a ligand on the development and/or differentiation of cells, tissues, organs and organisms are contemplated. Non-limiting examples include incubating a ligand with either major histocompatibility complex (MHC) class II-negative cells or single pluripotent myeloid-lymphoid initiating cells (ML-IC) and assessing cell fate by cytological and immunological techniques according to either Inaba E. *et al.*, 1993, PNAS 90:3038 or Punzel M *et al.*, 1999, Blood 93:3750.

6.2. EXAMPLE 2: DIABETES

Peripheral insulin resistance is the major pathogenic mechanism which causes type II diabetes, the fourth leading cause of death by disease and is the leading cause of blindness, renal failure and amputation. Insulin stimulates glucose uptake in muscle and fat cells, glycogen synthesis in liver and muscle cells and fat synthesis in fat and liver cells and the inhibition of glucose production in liver cells. NIDDM is characterized by impaired insulin-stimulated glucose uptake into skeletal muscle and adipocytes, impaired inhibition of liver gluconeogenesis and potentially misregulated insulin secretion. The pathway is only partially understood and the molecules responsible for peripheral insulin resistance are not known making it amenable to the methods of the instant invention.

Insulin binds to the α subunit of its dimeric receptor inducing the receptor's cytosolic β subunit tyrosine kinase activity to phosphorylate itself and nearby proteins. Insulin triggers activation of DNA and protein synthesis, activation of anabolic metabolic pathways and inhibition of catabolic metabolic

WO 02/058533

PCT/US01/43348

pathways. A series of proteins IRS-1, IRS-2, IRS-3, IRS-4, Gab-1 and p62 dock proteins all can bind the phosphorylated insulin receptor and can be substrates for it. IRS-1 appears to be most involved with the receptor but all of these are activators of phosphatidylinositol 3 kinase, which causes the transport of the striated muscle/adipose tissue specific glucose transporter GLUT 4 from the golgi in the cytoplasm to the plasma membrane where it transports glucose which is then phosphorylated by hexokinase. (Glut 2 is present on liver and β cells of pancreas). Insulin also up regulates glycogen synthase which catalyzes the final step of the conversion of glucose into glycogen but it is believed that the defect occurs in the first half of this signaling pathway.

The liver and the muscle account for most of the glucose metabolized and hence cells from these organs will be used in these studies. Diabetic patient muscle biopsies may be challenged with insulin and/or gliclazides as may be muscle biopsies from healthy individuals. The individuals may be relatives of the patients, some of whom have no overt symptoms of diabetes and a completely normal response to insulin. Defects in insulin action precede overt disease and are seen in nondiabetic relatives of diabetic patients. Differential display cDNA libraries may be prepared from diabetic patients and healthy individuals. A second differential display cDNA libraries may be prepared from patient biopsies challenged with insulin and/or gliclazides and biopsies from healthy patients. These cDNA libraries may then be expressed as proteins. Ligands which bind the expressed proteins may be isolated using the methods described in the invention (e.g., HPLC/ mass spectroscopy).

The ligands may be assayed for the effect on glucose uptake following insulin stimulation. 3T3-L1 adipocyte and L6 myocyte cell lines (ATOC) may be used as cell models for glucose metabolism. From 2×10^5 to 1×10^{10} cells may be plated in each well of a 96 well plate and medium containing a known concentration of glucose and 1 μ M to 10 μ M of each ligand is added to wells in triplicate. Minimally, a negative (no insulin, no ligands) and a positive (insulin, no ligands) control are performed. Insulin is next added to the wells at a low and a high concentration. After 2 hours incubation in a CO_2 incubator, glucose levels may be determined using a glucose meter. The ligands which affected

WO 02/058533

PCT/US01/43348

glucose metabolism following insulin stimulation in the cell lines may then be tested using the same assay with fresh skeletal muscle and adipose tissue biopsy from Type II diabetic patients. Cells suspended from the tissue biopsy may be plated at the same density in wells of a 96 well plate and the same assay as above repeated with each sample in duplicate. If the ligands decreased peripheral insulin resistance in these tissue biopsies, the ligand gene combination may represent a validated target in the treatment of peripheral insulin resistance which may be tested further and mapped in the metabolic signaling pathway of insulin.

6.3. IDENTIFICATION OF TARGETS IN MOLECULAR PATHWAYS OF KNOWN GENES

The approach used above may be used to identify and determine the function of unknown genes within the signaling pathways of pluripotent secreted proteins and to isolate the therapeutic effect from the toxic effect in a tissue specific way. TGF β 1 is a well known potent growth inhibitor in many cell types and the type II TGF β receptor, Smad 2 or Smad 4 are known to be mutated in a number of cancers (Kim SJ, 2000, Cytokine Growth Factor Rev. 11: 159). Some tumor suppressor genes (DPC4) are members of this SMAD family and are potent down regulators of T cell immune responses (Prud'homme GJ, 2000, J. Autoimmun. 14:23). Modulation of this growth inhibition and apoptosis induction pathway may be used to develop novel therapies to inhibit cancer cell growth, induce tolerance of T cells in autoimmunity and break tolerance to cancer antigens by blockade of this TGF β pathway.

One of the limiting factors has been that TGF β 1 also induces deposit of the extracellular matrix including up regulation of fibronectin, collagen, plasminogen activator inhibitor-1 and tissue inhibitors of matrix metalloproteases while down regulating matrix degrading proteases such as interstitial collagenase. Massague, 1990, J Ann Rev Biochem 6:597. Overproduction of matrix components is the major finding in tissue fibrosis an important cause of end stage renal and other diseases (Bloch GC, 2000, NEJM 342:1350). Decreased fibronectin production is often observed in cancer causing decreased

WO 02/058533

PCT/US01/43348

cellular adhesion and increased metastasis (Kombhrit *et al.*, 1996, *FASEB J* 10:248). TGF β induces these effects on ECM through a Smad independent pathway in which c-jun N-terminal kinase (JNK, a member of the MAP kinase family) activated to modulate cJUN (member of the AP-1 family of transcription factors) and ATF 2 (another transcription factor) (Hocavar *et al.*, 1999, *EMBO J* 18:1345). The pleiotropic effects of TGF β may be dissected out by targeting jun and smad pathways separately. To this end, primary human T cells and fibroblasts may be split into two and half of the cells may be transfected with a retroviral vector containing antisense jun or SMAD. Alternatively this may be achieved with a different vector or the cells may be transduced with a peptide reactive with either smad or jun. The resulting cell lines may then be stimulated with TGF β and cDNA's may be cloned which may be differentially expressed between stimulated and unstimulated cells and then cells with either pathway blocked using microarray analysis or other techniques of differential expression. Once cDNAs have been identified the expression of which is only associated with one of the pathways (but the function of which is unknown), these cDNAs can then be expressed as proteins, ligands binding to them can be isolated using the biochemical binding assay and resolution by HPLC and mass spectroscopy. The ligands can then be tested for the ability to block or induce either proliferation (in a PCNA based assay as described above) or secretion of the extracellular matrix. The extracellular matrix assay would measure fibronectin deposition, a major component of the extracellular matrix over a 48 hour period in a 96 well plate using an ELISA assay for fibronectin. In this way, genes can be identified and targets can be validated which are associated with the antiproliferative effect of the protein but not the profibrotic effect and visa versa. A similar approach may be used to look at any stimulus to a cells or tissue to identify new members of the molecular pathway and validate them as drug targets.

30

WO 02/058533

PCT/US01/43348

7.1. PHENOTYPE TO GENOTYPE

7.1.1. PHENOTYPE DETECTION

Tumor cell apoptosis and proliferation assays described in Sections 6.1.3.1 and 6.1.3.2. may be adapted to high throughput screening using, for example, a 384 well plate format (Applied Biosystems FMAT 8100). Apoptosis and necrosis may be assayed simultaneously. For apoptosis and necrosis the Cy5.5 Annexin V assay and TOTO 3 reagents respectively may be used (Applied Biosystems). Cy5.5 labeled anti-PCNA antibody (PC-10, Santa Cruz Biotechnology) may be used to assay cell proliferation. Non-limiting examples of human breast cancer cell lines which may be assayed include: MCF-7, NCI/ADR H5578T, MDA-MB-22231/ATCC, MDA-MB-4355, MDA-N, BT-549, T-47D (NCI, ATCC). Non-limiting examples of human prostate cancer cell lines which may be assayed include: DU-145, PC-3, LNCaP. Non-limiting examples of human colon cancer cell lines which may be assayed include: COLO 205, HCC-2998, HCT-15, HCT-116, HT29, KM12, SW-620. Non-limiting examples of human lung cancer cell lines which may be assayed include: A549/ATCC, EKVX, HOP-62, HOP-92, NCI-H23, NCI-H226, NCI-H322M, NCI-H460, NCI-H522. From 1×10^5 to 1×10^8 cells may be plated in each well of a 384 well plate. Medium containing 1 μ M to 1 M and preferably 1 μ M to 10 μ M of each potential ligand in a ligand library (non-limiting examples of which are listed in section 5.1.2 above) is added to wells are tested in triplicate. Negative (no ligands) and positive (staurosporine) controls are included. The ligands having the phenotypic effect at a concentration of $\leq 20 \mu$ M are good candidates for target identification according to the invention.

7.1.2. TARGET IDENTIFICATION

An important advantage of the invention is that, unlike the prior art, the target of a ligand which is found to have an effect in one or more bioassays, may be identified using the ligand. There are a number of approaches which may be used to identify the target according to the invention.

In a first series of embodiments, a potential target is a protein displayed on the surface of a cell. According to one non-limiting example, a full length

WO 02/058533

PCT/US01/43348

human cDNA library is expressed in the pDisplay vector (Invitrogen). This vector targets the protein to and anchors it in the cell membrane on the surface of eukaryotic cells. In another non-limiting embodiment of the invention, a full length human cDNA library is expressed in the pYTI1 yeast display vector or similar vector transfected into the EBY100 *Saccharomyces cerevisiae* strain (Invitrogen). In still another non-limiting embodiment of the invention, a full length human cDNA library is expressed on the surface of insect cells using baculovirus vector (Ernst W et al. 1998, Nucleic Acids Research 26:17-8). These systems allow full length proteins to be expressed on the surface as opposed to prokaryotic systems which only allow peptides to be expressed.

In alternative embodiments, a polynucleotide library can be expressed as a peptide alone or a fusion on the surface of a cell or a virus (e.g., bacteriophage, T7, or M13). Non-limiting examples include a polynucleotide library generated from human or infectious agent. In a specific embodiment of the invention, a cDNA library is expressed as dodecapeptides in the pFliTx vector (Invitrogen) or similar. According to this embodiment when the vector is expressed in *E. coli*, the peptide is displayed in the active site loop of the thioredoxin protein and inside the bacterial flagellin gene. In another embodiment of the invention, potential targets may be displayed as peptides on a ribosome display system in which the peptide is fused to the RNA encoding it by treatment with puromycin (Roberts RW et al., 1977, PNAS 94:12297). All other display systems (including but not limited to retrovirus, adenovirus) may be used in accordance with the invention to display cDNAs or peptides.

7.1.3. SEPARATION

Potential targets displayed by any of the above methods may be exposed to the ligand. The ligand may be either immobilized on a surface, bead or column or it may be in solution depending on the separation method to be used. In a first embodiment of the invention, the ligands may be directly immobilized on the surface, directly labeled or detected. In a second embodiment of the invention, the ligands may be derivatized with an affinity label to facilitate collection of the ligand-target pair where the target is displayed as illustrated in

WO 02/058533

PCT/US01/43348

the foregoing examples. Non-limiting examples of such affinity labels include biotin, digoxigenin, or an antibody. Displayed targets which bind the ligand may then be separated from those which do not bind and the sequence encoding the target is identified by standard cloning and DNA sequencing techniques.

5 In a first embodiment of the invention, cells can be "stained" with fluorescently labeled or biotinylated ligand (the latter combined with FITC-avidin) and sorted using a flow cytometer (MoFlo HTS Cytometer, Becton Dickinson FACS) into wells of a plate, a tube, etc. The cells may then be grown using standard cell culture techniques. According to a first non-limiting
10 example, the gene encoding the drug's receptor may then be cloned by plasmid recovery from COS 1 cells by using the effect of the large T antigen effect on the SV40 origin of replication. According to a second non-limiting example, PCR may be used to recover the plasmid insert.

In a second embodiment of the invention, cells, viral particles or peptide-nucleotide fusions may be selected using drug coated magnetic beads, a drug
15 coated surface (e.g., a well for panning) or a drug coated column. A high density of drug ligands on the surface, beads or column is desirable to increase the avidity of low affinity interactions. The drug may be attached to the surface, beads or column via an affinity label (e.g., avidin, digoxigenin) and elution may
20 be achieved after one or more washing steps. In the case of magnetic beads, magnets may then be used to isolate beads during the wash to recover bound cells, viral particles or peptide-nucleotide fusions. In the case of panning, the supernatant is poured off after each successive washing step with the cells, viral particles or peptide-nucleotide fusions retained in the wells. Elution from a
25 column may be achieved by standard techniques. In the case where the ligands were derivatized with an affinity label, cells, viral particles or peptide-nucleotide fusions may be eluted from the column by applying excess free affinity label to the column.

Once separated, target expressing cells or viral particles can be grown as
30 appropriate. Then the cDNA encoding the target may be recovered by standard molecular biology techniques (e.g., plasmid recovery or PCR). In the case of purified peptide-nucleotide complexes, the partial cDNA sequence would be

WO 01/058533

PCT/US01/43348

Identified using RT PCR. Using the above approach the target can be purified and cloned using one or more rounds of selection. In this way, the DNA sequence encoding a previously unknown drug target can be isolated and used to clone the cDNA encoding the drug target.

Once the cDNA encoding the drug target has been identified, the cDNA can be used to study differential expression in cells from disease tissues as in section 6.1. If the target is differentially expressed between disease and normal cells, specificity is established and the ligands interacting with that target may be tested *in vitro* and *in vivo* bioassays for that disease.

Thus the target associated with a function in the phenotypic assay is identified employing the invention.

7.2. TARGET IDENTIFICATION BY PROTEOMICS

Target identification may also be achieved by adapting the method set forth in section 6.1.2. to combine the ligand of interest with one a plurality of potential targets, collecting ligand target pairs, and optimally dissociating the ligand and target. Subsequently, the target may be identified. In one embodiment of the invention, the target is a protein which may be identified by common techniques (e.g., amino acid sequencing, mass spectroscopy and/or NMR). Once the protein has been identified, its association with diseased cells may be determined using standard proteomics techniques.

8.1. MAPPING SIGNALING PATHWAYS

Once a number of genes have been shown to be involved in a particular molecular pathway of disease pathogenesis, a targeted component can be mapped within the molecular pathway relative to other molecular pathway components. Ligands which bind to different molecular pathway components may be derivatized with photoactivatable crosslinkers. At least one of the known molecular pathway components is fused with a marker such as GFP. Then the following may be combined *in vivo* or *in vitro*: (i) a derivatized ligand which binds the known molecular pathway component, (ii) the marked pathway component, e.g., GFP fusion protein, (iii) at least one derivatized ligand which

WO 02/058533

PCT/US01/43348

binds or may bind another molecular pathway component, and (iv) other molecular pathway components. The crosslinking stimulus is applied and each component of the resulting complex is identified. In this way each molecular pathway components may be mapped relative to other components with which it interacts. A further advantage of the invention is that pathway effectors may be identified by this method. In addition, the profile of each pathway component may be compared with known drugs acting via that pathway, if any, and comparative studies can be done in cell based assays of different diseases caused by that pathogenic pathway. This information can be used to validate and select the best target for a given disease indication. As an alternative, this information may be used to select the best therapies for a particular patient using pharmacogenetics.

9.1. LEAD OPTIMIZATION

Because a large number of chemical leads may be characterized at the biochemical and phenotypic levels, a structure activity relationship (SAR) may be established to serve as a basis for lead optimization. If a few molecules with similar activities are identified, the SAR can be determined by comparing their structures with activity in the assays. The target directed synthesis technology can be employed to crosslink molecules binding close to each other indicating if their activity is mediated through the same active subsite on the protein or through different subsites on the protein target. In this way additional different functional subsites on the target can be mapped and different mechanisms can be interpreted from the phenotypic findings with molecules binding to those subsites (*e.g.*, agonist vs. antagonist).

The second use of target directed synthesis is to increase the affinity of a ligand for its target and thus make the ligand more useful to link phenotype to genotype as well as making a better drug lead. Photoactivatable crosslinkers on one of the functional groups of the ligand scaffold may be used to link ligands bound to the target thus using the target molecule as a template. This linking should increase the affinity of binding to the target by at least 2- to 10- fold and

WO 02/058533

PCT/US01/43348

further enhance the structural diversity of the library in a target directed and biologically relevant way.

10. IN SILICO APPROACH TO LINKING PHENOTYPE WITH

5 GENOTYPE

The instant invention provides a method to establish a chemical fingerprint of ligand-target (genotype) and ligand-bioassay (phenotype) for each ligand or set of ligands which can be matched in silico to associate phenotype with genotype.

10 The present invention provides a first information retrieval system wherein ligand-target pairing experimental data will be stored. The present invention provides a second information retrieval system wherein the effects of each ligand in each bioassay tested will be stored. The present invention provides a third information retrieval system wherein the function and/or the expression pattern of each target, if known, will be stored. These systems may
15 be optionally integrated to facilitate use.

In one embodiment of the invention, data entered into the systems may be obtained by a shotgun approach wherein all targets are tested for binding to ligands or all ligands are tested in each bioassay. For example, the set of targets
20 may encompass up to all expression products of up to and including all genes in the genome of a selected organism. Each target is then used to screen a library of ligands to identify ligands which bind. This data is entered into the first information retrieval system.

According to another example, the effect of each member of a large
25 combinatorial chemical library of ligands may be assayed in each available bioassay. This data is entered into the second information retrieval system.

In another embodiment of the invention, data entered into the system is obtained by a focused analysis of ligands which bind selected targets in a specific disease or the phenotype induced by selected ligands in selected
30 bioassays. This data is entered into the first or second information retrieval system as appropriate.

WO 01/058533

PCT/US01/43348

These systems may then be used to guide the user in predicting target function even in the absence of differential expression data or a particular disease focus. In addition, these systems may guide the user in selecting ligands and targets with specific effects. A further advantage is that this system may reduce the number of binding experiments and bioassays necessary. Other advantages will be apparent to one skilled in the art.

In one embodiment of the invention, a user selects a target of interest. Next, the user identifies ligand(s) which bind the target of interest either experimentally or from the first information retrieval system. The user then queries the second information retrieval system with the identified ligand(s) to determine the phenotype(s) associated with each ligand. In this way, a target may be associated with one or more phenotypes.

In another embodiment of the invention, a user selects a phenotype of interest. Next, the user identifies ligand(s) which modulate the selected phenotype either experimentally or from the second information retrieval system. The user then queries the first information retrieval system with the identified ligand(s) to identify target(s) to which the ligand(s) binds. In this way, a phenotype may be associated with one or more targets.

In a another embodiment of the invention, these information retrieval systems may be combined with target functional information and/or expression analysis data to guide the user in validating targets and drug leads. In a first example of this embodiment, a user may choose targets X and Y which are proteins. The user obtains expression data which indicates that the gene encoding X is expressed in normal cells but is not expressed in tumor cells. The user obtains further expression data which indicates that the gene encoding Y is not expressed in normal cells but is expressed in tumor cells. The user then queries the first information retrieval system. The results of this query are shown in Table 2.

WO 02/058533

PCT/US01/43348

Table 2.

Target	Ligands that Bind
X	1
X	2
X	3
Y	2
Y	3
Y	4

The user then queries the second information retrieval system. The results of this query are shown in Table 3.

5

Table 3.

Ligands	Phenotype
1, 2, 3	Angiogenesis
2, 3, 4	Proliferation

WO 02/058533

PCT/US01/43348

According to this example, the user may select target Y as a valid target for cancer therapy and may select ligand 4 for its ability to specifically bind Y and not X. Thus, the invention is able to guide the user in validating targets and identifying drug leads.

5 In a second example of this embodiment, the phenotype to genotype approach has been used to determine that ligands 1, 2, and 3 induce apoptosis in a bioassay; ligands 3, 4, and 5 stimulate angiogenesis; and ligands 1, 3, and 6 induce necrosis. This information is stored in an information retrieval system. In a high throughput binding assay, it is discovered that ligands 3 and 4 bind to
10 target X with $K_d < 50 \mu M$. A search of the information retrieval system will indicate to one skilled in the art that (i) target X may be involved in angiogenesis, (ii) ligand 3 is a poor candidate for a drug lead, and (iii) ligand 4 may be a good candidate for a drug lead.

15 11. AUTOMATION OF THE METHODS OF THE INVENTION

A highly automated approach such as those shown diagrammatically in Figs. 18 and 19 is another embodiment of the present invention. This includes high throughput expression vector construction, protein production, and purification facility capable of producing >20 proteins a week in sufficient
20 amounts to determine ligands from a compound library. This is followed by the use of a high throughput assay such as the Chemical Array Assay to identify scaffold target pairs. These scaffold target pairs comprise the chemical array database which has the uses outlined in Fig. 17.

For high throughput expression vector construction, a cDNA encoding
25 one of the proteins in the human proteome from, for example, NCBI, Stratagene, or Incyte is inserted into a DES expression vector (Invitrogen) using an automated fluid handling system (Tecan) in a 96 well format. The DES expression vector adds a secretion signal and a his-tag to the encoded protein so that it is secreted into the media and can be purified using a nickel column that
30 binds the his-tag. The vectors are then transfected into competent *E. coli* cells, and the cells are propagated. The expression vector can be extracted from the *E. coli* cells using a robotic fluid handler to add a standard lysis reagent to lyse the

WO 02/058533

PCT/US01/43348

cells and to apply the lysate to Qiagen columns to purify the expression vector. In a particular embodiment, the lysate is purified using the QIAwell 96 Ultra Plasmid Kit which uses a Qiafilter 96 well plate for lysate clearing, QIAwell 96 well plates for purification of the plasmid DNA, and QIAprep 96 well plates for desalting each plate sequentially on the QIAvac 96 automated vacuum device. If desired, cells containing the expression vector with the cDNA insert in the proper reading frame are selected using standard methods. For example, the expression vector can be restriction enzyme digested or sequenced to determine whether it contains the cDNA insert in-frame.

The expression vector containing the insert is then transfected into *Drosophila* S2 cells (Invitrogen) using standard calcium phosphate transfection methods and grown in *drosophila* expression media (Invitrogen) in 6-12 flasks per vector in the Select automated tissue culture system (Automation Partnership). Each Select system can handle up to 150 flasks or up to 40 separate cell lines expressing different proteins, and using multiple Select's in parallel can increase throughput to 600 proteins per week. After 24 hours, copper sulfate is added to the medium to induce protein expression and on day 3 and 7 the supernatant is collected and passed through the nickel column in 96 well format (Qiagen QIAexpress protein purification system) on a Biorobot (Qiagen). A Tecan fluid handler then transfers an aliquot of this protein to PHAST gel (Pharmacia) for SDS analysis or other quality control analysis (QC).

The rest of the sample is transferred by the reagent storage retrieval system (Haystack) to the Chemical Array Array (e.g., in any of the assay methods described herein) and to the freezer for storage. For example, a robotic fluid handler (Tecan) can be used to combine the purified protein target with a library of candidate ligands to allow one or more of the candidate ligands to bind the target protein in the wells of a 96 well plate. This 96 well plate can then be transferred to an HPLC (Waters 2790) which can inject the assay mixture containing the target protein and candidate ligands from 96 well plates and run up to 6 columns in parallel for the isolation of the target protein with bound ligands. The fraction containing the target with bound ligand can be collected using a fraction collector (Gibson). In an alternative embodiment, a robotic fluid

WO 01058533

PCT/US01/43348

handler (Tecan) is used to combine the purified protein target with a library of candidate ligands to allow one or more of the candidate ligands to bind the target protein in the wells of a 96 well plate. This 96 well plate contains, for example, cartridges with a resin capable of separating target proteins from unbound ligands to isolate the target protein with bound ligands into a second 96 well plate upon evacuation by a robot (Tecan or Qiagen). In an alternative embodiment, the binding occurs in a 96 well plate, and then a fluid handler (Tecan) transfers the sample to a second 96 well plate including the cartridges for separation. In still another embodiment, the cartridges are spin columns which are available in multiwell formats (Pharmacia). Chip based and capillary LC based separations can also be used. A detergent or other denaturant can be added by the fluid handler (Tecan) to release the bound ligands from the protein, and then the released ligands are added to an appropriate instrument for analysis. For example, the ligands can be injected into a mass spectrometer using a reverse phase column on an HPLC containing an autoinjector (Waters), spotted on a filter for MALDITOF mass spectrometry analysis, or applied to an NMR, IR, FTIR, or UV spectrometer. In an alternative embodiment, the target protein with bound ligands is loaded or spotted onto the 96 well format MALDITOF (Bruker Daltonics) using a fluid handler (Tecan). In another alternative embodiment, the target protein with bound ligands is evacuated onto a filter (for example, nitrocellulose) in a 96 well format by evacuation with a robot (Tecan). In another embodiment, the evacuation onto this same filter is performed in the same step as the evacuation of the 96 well cartridges by placing the filter between the cartridges and the vacuum device. The MALDITOF then dissociates the target protein and ligands from each of the 96 spots and generates a mass spectrum for the compound and/or complex. After data processing by the information systems described herein, the identity of the ligand and its target are entered into the Chemical Array Database. Any of these methods can be performed in 384, 1536 well, chip based, or other formats. Similarly, any of the data can be entered and managed using a laboratory information management system (LIMS) based on IDBS Activity Base or Price Waterhouse, or other LIMS software/systems.

WO 02/058533

PCT/US01/43348

Similar methods can be applied for other transient expression based production systems including, but not limited to, HEK293 cells, CHO, or COS cells. Alternatively, other automated or semi-automated production systems can be used, such as roller bottle systems, Stir tank systems (e.g., Celligen Plus from New Brunswick), or capillary cell culture systems (Amicon). In another embodiment, a semiautomated process, such as a 1 L or larger bioreactor from New Brunswick, is used to grow cells such as HEK293 cells (Life Technologies) transiently transfected with expression constructs constructed as described above based upon the pCDNA family of vectors (Invitrogen). Transiently transfected CHO cells can also be used. The transfection in these cell types can be efficiently achieved using Lipofectamine 2000 (Life Technologies). In alternative embodiments, other transfection strategies are used (for example, electroporation, Calcium Phosphate, Lipofectin, Lipofectamine Plus (Life Technologies), or other standard techniques). These cells are grown in DMEM or in other standard mediums with serum or in serum free forms using standard methods. In addition, alternative expression vectors, such as those appropriate for the various cell lines mentioned as indicated in the catalogue of Invitrogen, other vector companies, the scientific literature, or those which would be apparent to those skilled in the art.

If desired, a clone selection step can be performed, resulting in stable producer cell line based production systems (e.g., CHO or *E. coli* based systems). Exemplary clone selection steps include growing the cells in the presence of an selective antibiotic, e.g., G418, in a multi-well format to select cells likely to contain the expression vector, and then checking each well for the presence of the secreted protein using a standard ELISA assay or other standard assay to detect the his-tag present in the protein.

In addition, high throughput production and screening techniques can be used for any of the methods of the invention. For example, any binding assay (chip, filter, radiolabelled, fluorescent, surface plasmon resonance, etc.), production method (e.g., mammalian cells such as CHO, HEK 293, Cns; insect cells such as *Drosophila*, bacteria such as *E. coli*, or yeast such as *Pichia*), production systems (e.g., bioreactors (New Brunswick systems by Brandel, flask

WO 02/058533

PCT/US01/43348

based, cell cube, surface bound, suspension cultures, serum containing media, or serum free media), and any purification method (HIS tag/nickel column, GST/glutathione, immo, or other affinity column) can be used. Any of these automated and/or high throughput methods can be performed with multiple systems acting in parallel, such as multiple robotic systems (such as multiple 5 SelecT robots from Automation Partnership). For example, 2, 2, 4, 5, 6, 8, 10, 10^2 , 10^3 , 10^4 , 10^5 , 10^6 , or more targets can be assayed in parallel to select ligands that bind the targets. Similarly, 2, 5, 10, 10^2 , 10^3 , 10^4 , 10^5 , 10^6 , 10^7 , 10^8 , or 10^9 or more small molecules of interest can be assayed in parallel to select target 10 molecules that bind the small molecules.

Other Embodiments

From the foregoing description, it will be apparent that variations and modifications may be made to the invention described herein to adapt it to 15 various usages and conditions. Such embodiments are also within the scope of the following claims.

Various publications and patent applications are cited herein, the contents of which are hereby incorporated by reference in their entireties to the same extent as if each independent publication or patent application was specifically 20 and individually indicated to be incorporated by reference.

WO 01/058533

PCT/US91/43348

CLAIMS

1. A method for selecting a candidate ligand which binds a target molecule, said method comprising:
- 5 (a) contacting an *in vitro* sample comprising a target molecule with a library of candidate ligands under conditions that allow complex formation between said target molecule and one or more said candidate ligands, wherein said library comprises at least two different chemical scaffolds or comprises at least 11 different compounds;
- 10 (b) isolating said complex;
- (c) recovering one or more said candidate ligands from said complex; and
- (d) identifying one or more recovered candidate ligands.
2. The method of claim 1, wherein step (d) comprises determining the
- 15 MS, IR, FTIR, NMR, and/or UV spectrum of said recovered candidate ligand.
3. The method of claim 1, wherein at least 100 different candidate ligands are simultaneously contacted with said target molecule.
- 20 4. A method for selecting a candidate ligand which binds a target molecule, said method comprising:
- (a) contacting an *in vitro* sample comprising a target molecule with a library of candidate ligands under conditions that allow complex formation between said target molecule and one or more said candidate ligands;
- 25 (b) isolating said complex;
- (c) recovering one or more said candidate ligands from said complex; and
- (d) determining the mass to charge ratio of an isotope or fragment peak in the mass spectrum of a recovered candidate ligand, thereby identifying said recovered candidate ligand.
- 30 5. The method of claim 4, wherein at least 100 different candidate ligands are simultaneously contacted with said target molecule.

WO 02/058533

PCT/US01/43349

6. The method of claim 4, wherein step (d) further comprises determining the mass to charge ratio of the parent peak in the mass spectrum of said recovered candidate ligand.
- 5
7. A method for selecting a candidate ligand which binds a target molecule, said method comprising:
- (a) contacting an *in vitro* sample comprising a target molecule of unknown biological function with a library of candidate ligands under conditions that allow complex formation between said target molecule and one or more said candidate ligands;
- 10 (b) isolating said complex;
- (c) recovering one or more said candidate ligands from said complex; and
- (d) determining the MS, IR, FTIR, NMR, and/or UV spectrum of a recovered candidate ligand, thereby identifying said recovered candidate ligand.
- 15
8. The method of claim 7, wherein at least 100 different candidate ligands are simultaneously contacted with said target molecule.
- 20
9. A method for selecting a candidate ligand which binds a target molecule, said method comprising:
- (a) contacting an *in vitro* sample comprising a target molecule with one or more candidate ligands under conditions that allow complex formation between said target molecule and one or more said candidate ligands;
- 25 (b) isolating said complex;
- (c) recovering one or more said candidate ligands from said complex; and
- (d) determining the IR, FTIR, NMR, and/or UV spectrum of a recovered candidate ligand, thereby identifying said recovered candidate ligand.
- 30
10. The method of claim 9, wherein at least 100 different candidate ligands are simultaneously contacted with said target molecule.

WO 02/058533

PCT/US01/43348

17. A method for selecting a candidate ligand which binds a target molecule, said method comprising:

- (a) contacting an *in vitro* sample comprising a first target molecule and a second target molecule with a library of candidate ligands under conditions that allow complex formation between said first target molecule and one or more said candidate ligands and allow complex formation between said second target molecule and one or more said candidate ligands;
- (b) isolating a first complex comprising said first target molecule bound to a candidate ligand and isolating a second complex comprising said second target molecule bound to a candidate ligand;
- (c) recovering one or more said candidate ligands from said first complex and/or from said second complex; and
- (d) identifying one or more recovered candidate ligands.

12. The method of claim 11, further comprising contacting said sample with a competitor ligand known to bind said target molecule, said first target molecule, or said second target molecule.

13. A method for determining the biological function of a target molecule, said method comprising:

- (a) contacting an *in vitro* sample comprising a target molecule of unknown biological function with a library of candidate ligands under conditions that allow one or more said candidate ligands to bind said target molecule;
- (b) selecting a candidate ligand which binds said target molecule; and
- (c) measuring the effect of said selected candidate ligand in a biological assay, thereby determining the biological function of said target molecule.

14. The method of claim 13, further comprising identifying said selected candidate ligand.

WO 02/058533

PCT/US01/43348

15. A method for determining the biological function of a target molecule, said method comprising:

- (a) contacting an *in vitro* sample comprising a target molecule that is upregulated or downregulated in a disease state, in the presence of a physiological stimulus, or during a specific cellular or biological process with a library of candidate ligands under conditions that allow one or more said candidate ligands to bind said target molecule;
- (b) selecting a candidate ligand which binds said target molecule; and
- (c) measuring the effect of said selected candidate ligand in a biological assay, thereby determining the biological function of said target molecule.

16. The method of claim 15, further comprising identifying said selected candidate ligand.

17. The method of claim 15, wherein said selected candidate ligand increases the activity of said target molecule in said biological assay.

18. The method of claim 15, wherein said selected candidate ligand decreases the activity of said target molecule in said biological assay.

19. A method for determining the biological function of a target molecule, said method comprising:

- (a) contacting an *in vitro* sample comprising a target molecule with a library of candidate ligands under conditions that allow one or more said candidate ligands to bind said target molecule;
- (b) selecting a candidate ligand which binds said target molecule; and
- (c) measuring the effect of said selected candidate ligand on a tissue from a organism having a disease or disorder or undergoing a specific cellular or biological process in the presence or absence of a physiological stimulus, thereby determining the biological function of said target molecule.

WO 02/65853

PCT/US91/43348

20. The method of claim 19, wherein said tissue is human tissue.

21. A method for reacting two ligands that bind a target molecule of interest, said method comprising contacting a cell or *in vitro* sample comprising
5 a target molecule of unknown secondary or tertiary structure with a first ligand comprising a first crosslinker and with a second ligand under conditions that allow said target molecule to bind said first ligand and said second ligand and allow said first crosslinker to covalently bind said second ligand, thereby
10 generating a crosslinked product comprising said first ligand and said second ligand.

22. A method for reacting two ligands that bind a target molecule of interest, said method comprising contacting a cell or *in vitro* sample comprising
15 a target molecule with a first ligand comprising a first crosslinker and with a second ligand, wherein the location or the tertiary structure of the binding site in said target molecule for said first ligand or said second ligand is unknown, and wherein said contacting is conducted under conditions that allow said target molecule to bind said first ligand and said second ligand and allow said first crosslinker to covalently bind said second ligand, thereby generating a
20 crosslinked product comprising said first ligand and said second ligand.

23. A method for reacting two ligands that bind a target molecule of interest, said method comprising contacting a cell or *in vitro* sample comprising
25 a target molecule with a first ligand comprising a first crosslinker and with a second ligand, wherein said contacting is conducted under conditions that allow said target molecule to bind said first ligand and said second ligand and allow said first crosslinker to covalently bind said second ligand, thereby generating a crosslinked product comprising said first ligand and said second ligand that has an affinity for said target molecule that is greater than the affinity of said first
30 ligand or said second ligand for said target molecule.

WO 02/058533

PCT/US01/43348

24. A method for reacting two ligands that bind different target molecules, said method comprising contacting a cell or *in vitro* sample comprising a first target molecule and a second target molecule with a first ligand comprising a first crosslinker and with a second ligand, wherein said
 5 contacting is conducted under conditions that allow
 (i) said first protein to bind said first ligand,
 (ii) said second protein to bind said second ligand, and
 (iii) said first crosslinker to covalently bind said second ligand, thereby generating a crosslinked product comprising said first ligand and said second
 19 ligand;
 and wherein the location or the tertiary structure of the binding site in said first target molecule for said first ligand and/or the location or the tertiary structure of the binding site in said second target molecule for said second ligand is unknown.

15 25. The method of claim 24, wherein the generation of said crosslinked product indicates that said first protein and said second protein interact *in vivo*.

26. A method for isolating a second protein which binds a first protein, said method comprising:
 29

(a) contacting a cell or an *in vitro* sample comprising a first protein and a second protein with a first ligand comprising a first crosslinker and with a second ligand under conditions that allow

(i) said first protein to bind said first ligand,
 25 (ii) said second protein to bind said second ligand, and
 (iii) said first crosslinker to covalently bind said second ligand, thereby generating a crosslinked product comprising said first ligand and said second ligand and generating a complex comprising said crosslinked product, said first protein, and said second protein;
 30 (b) isolating said complex; and
 (c) identifying said first protein and/or said second protein in said complex or recovered from said complex.

WO 02058533

PCT/US01/43348

27. The method of claim 26, wherein said first protein comprises a detectable group.

5 28. The method of claim 26, wherein said second ligand comprises a crosslinker.

29. The method of claim 26, wherein the generation of said crosslinked product indicates that said first protein and said second protein interact *in vivo*.

10 30. The method of claim 26, wherein the affinity of said crosslinked product for said target molecule is greater than the affinity of said first ligand or said second ligand for said target molecule.

15 31. The method of claim 26, wherein said crosslinked product is used in drug discovery or development or lead optimization.

32. The method of claim 26, wherein said crosslinked product is used in the development of an agricultural or environmental agent.

20 33. A method for selecting a candidate target molecule which binds a small molecule of interest, said method comprising:
(a) contacting an *in vitro* sample comprising a small molecule of interest having a moiety other than an amino acid or having a molecular weight less than
25 4000 daltons with a library of candidate target molecules under conditions that allow complex formation between said small molecule of interest and one or more said candidate target molecules; wherein said target molecules are not expressed on the surface of phage;
(b) isolating said complex; and
30 (c) recovering one or more said candidate target molecules from said complex, thereby selecting one or more candidate target molecules which bind said small molecule of interest.

WO 02/058533

PCT/US01/43348

34. The method of claim 33, wherein, prior to step (a), said small molecule of interest is selected from a library of small molecules based on its effect in a biological assay.

35. A method for selecting a target protein which binds a small molecule of interest, said method comprising:

(a) expressing in a population of cells a protein fusion comprising a target protein covalently linked to surface protein, said expression being carried out under conditions that allow the display of said protein fusion on the surface of said cells;

(b) contacting said cells with a small molecule of interest having a moiety other than an amino acid or having a molecular weight less than 4000 daltons; and

(c) selecting said cells which bind said small molecule of interest, thereby selecting said target proteins which bind said small molecule of interest.

36. The method of claim 35, wherein said cell is a mammalian, bacterial, yeast, or insect cell.

37. A method for selecting a target protein which binds a small molecule of interest, said method comprising:

(a) expressing in a population of cells a protein fusion comprising a target protein covalently linked to surface protein, said expression being carried out under conditions that allow the display of said protein fusion on the surface of viruses released from said cells infected with said virus;

(b) contacting said viruses with a small molecule of interest, wherein said small molecule of interest

(i) is a nucleic acid,

(ii) is a carbohydrate,

(iii) is a lipid

(iv) has a moiety other than an amino acid.

WO 02/058533

PCT/US01/43348

(v) has a molecular weight less than 750 daltons, or

(vi) is not a molecule naturally produced by bacteria, and

(v) selecting said viruses which bind said small molecule of interest, thereby selecting said target proteins which bind said small molecule of interest.

5

38. The method of claim 37, wherein said virus is a bacteriophage or adenovirus.

39. A method for selecting a target protein which binds a small molecule of interest, said method comprising:

10

(a) expressing in a population of cells or an *in vitro* sample a library of target proteins, wherein each target protein is covalently linked to a nucleic acid encoding said target protein;

(b) contacting said cells or *in vitro* sample with a small molecule of interest having a moiety other than an amino acid or having a molecular weight less than 4000 daltons; and

15

(c) selecting said target proteins which bind said small molecule of interest.

40. The method of claim 39, further comprising identifying said selected target protein.

20

41. The method of claim 39, wherein at least 100 human target proteins are contacted with said small molecule of interest.

25

42. The method of claim 39, wherein said small molecule of interest is a non-naturally occurring molecule.

WO 01/058533

PCT/US01/43348

43. A method for selecting a candidate compound that binds or modulates the activity of a target molecule prior to validation of said target molecule as a drug target, said method comprising:

5 (a) contacting a cell or an *in vitro* sample comprising a target molecule that has not been previously validated as a drug target with a library of candidate compounds under conditions that allow one or more said candidate compounds to bind or modulate the activity of said target molecule; and

(b) selecting a candidate compound which binds or modulates the activity of said target molecule.

10

44. The method of claim 43, wherein said library comprises at least five candidate compounds.

45. The method of claim 43, further comprises the step of (c) measuring the effect of said selected candidate compound in a biological assay, thereby determining the biological function of said target molecule.

15

46. A method for selecting candidate compounds that bind or modulate the activity of target molecules, said method comprising:

20

(a) contacting a cell or an *in vitro* sample comprising a first target molecule and a second target molecule with a library of candidate compounds under conditions that allow one or more said candidate compound to bind or modulate the activity of said first target molecule and allow one or more said candidate compound to bind or modulate the activity of said second target

25

molecule;

(b) selecting a candidate compound which binds or modulates the activity of said first target molecule; and

(c) selecting a candidate compound which binds or modulates the activity of said second target molecule.

30

WO 02/058533

PCT/US01/43348

47. The method of claim 46, wherein said cell or *in vitro* sample comprises at least five target molecules, and wherein, for each of said target molecules, a candidate compound is selected that binds or modulates the activity of said target molecule.

5

48. An electronic database comprising at least 10 records of target molecules correlated to records of ligands and their ability to bind or modulate the activity of said target molecules.

10 49. The database of claim 48, comprising records for at least 0.5% of the proteins in the proteome of an organism.

50. An electronic database comprising at least 10 records of target molecule domains correlated to records of ligands and their ability to bind said domains.

15

51. An electronic database comprising a plurality of records of target molecules that have not been previously validated as drug targets correlated to records of ligands and their ability to bind or modulate the activity of said target molecules.

20

52. A computer comprising the database of claim 48, 50, or 51, and a user interface (i) capable of displaying one or more ligands that bind or modulate the activity of a target molecule whose record is stored in said computer or (ii) capable of displaying one or more target molecules that bind or have an activity that is modulated by a ligand whose record is stored in said computer.

25

53. An electronic database comprising at least 1000 records of compounds correlated to records of a phenotype in one or more biological assays effected by said compounds; wherein said biological assay involves a cell or *in vitro* sample that does not contain an exogenous copy of a nucleic acid encoding a protein that binds said compound.

30

WO 02/058533

PCT/US01/43348

54. A computer comprising the database of claim 53 and a user interface
(i) capable of displaying one or more phenotypes in one or more biological
assays for a compound whose record is stored in said computer or (ii) capable of
5 displaying one or more compounds that effects a phenotype whose record is
stored in said computer.

55. An electronic database comprising at least 10 records of target
molecules correlated to records of an expression profile or activity of said target
10 molecules.

56. An electronic database comprising a plurality of records of target
molecules that have not been previously validated as drug targets correlated to
records of an expression profile or activity of said target molecules.

57. A computer comprising the database of claim 55 or 56 and a user
interface (i) capable of displaying one or more expression profiles or activities of
a target molecule whose record is stored in said computer or (ii) capable of
displaying one or more target molecules that have an expression profile or
20 activity whose record is stored in said computer.

58. A method of identifying a target molecule associated with a
phenotype of interest, said method comprising:
(a) providing a first electronic database comprising a plurality of records
25 of phenotypes in a biological assay correlated to records of the ligands and their
ability to contribute to said phenotypes;
(b) receiving a selection of a phenotype of interest;
(c) identifying one or more ligands in said first database which cause said
phenotype of interest;
30 (d) providing a second electronic database comprising a plurality of
records of ligands correlated to records of the target molecules which bind said
ligands or have an activity that is modulated by said ligands; and

WO 02/058533

PCT/US01/43348

(e) identifying one or more target molecules in said second database that bind or are modulated by said ligand(s) which cause said phenotype of interest, thereby identifying one or more target molecules associated with said phenotype of interest.

5

59. The method of claim 58, wherein said phenotype of interest is associated with a disease state, and said target molecule is determined to promote or inhibit said disease state.

10

60. The method of claim 58 wherein said method is computer implemented.

61. A method of identifying a phenotype that is associated with a target molecule of interest, said method comprising:

15

(a) providing a first electronic database comprising a plurality of records of target molecules correlated to records of the ligands and their ability to bind or modulate the activity of said target molecules;

(b) receiving a selection of a target molecule of interest;

(c) identifying one or more ligands in said first database which bind or

20

modulate the activity of said target molecule of interest;

(d) providing a second electronic database comprising a plurality of records of ligands correlated to records of phenotypes in a biological assay caused by said ligands; and

(e) identifying one or more phenotypes in said second database caused by said ligand(s), thereby identifying one or more phenotypes associated with said target molecule of interest.

25

62. The method of claim 61, wherein said method is computer implemented.

30

WO 02/059533

PCT/US01/43348

63. A method of identifying a ligand that binds or modulates the activity of a target molecule of interest, said method comprising:

- (a) providing an electronic database comprising at least 10 records of target molecules correlated to records of the ligands and their ability to bind or modulate the activity of said target molecules;
- 5 (b) receiving a selection of a target molecule of interest; and
- (c) identifying one or more ligands in said database which bind or modulate the activity of said target molecule of interest.

10 64. The method of claim 63, wherein said ligand is used in drug discovery or development or lead optimization.

65. The method of claim 63, wherein said ligand is used in the development of an agricultural or environmental agent.

15 66. The method of claim 63, wherein said method is computer implemented.

20 67. The method of claim 63, further comprising comparing the chemical structures of two or more ligands which bind or modulate the activity of said target molecule of interest, thereby identifying functional groups in said ligands which promote the binding or modulation of said target molecule of interest.

25 68. The method of claim 63, further comprising comparing the chemical structures of two or more ligands which bind or modulate the activity of said target molecule of interest, thereby determining the frequency of one or more functional groups or scaffolds in the collection of said ligands.

WO 02/058533

PCT/US91/43348

69. The method of claim 63, further comprising generating one or more compounds that have one or more functional groups that are present in two or more of said ligands; wherein said compound is used in drug discovery or development or lead optimization.

5

70. A method of identifying a target molecule that binds or has an activity that is modulated by a ligand of interest, said method comprising:

- (a) providing an electronic database comprising at least 10 records of ligands correlated to records of the target molecules which bind or have an activity that is modulated by said ligands;
- (b) receiving a selection of a ligand of interest; and
- (c) identifying one or more target molecules in said database which bind or have an activity that is modulated by said ligand of interest.

10

71. The method of claim 70, wherein said method is computer implemented.

15

72. A method for determining the selectivity of a ligand of interest, said method comprising:

- (a) providing an electronic database comprising at least 10 records of target molecules correlated to records of the ligands and their ability to bind or modulate the activity of said target molecules;
- (b) receiving a selection of a ligand of interest; and
- (c) determining the number of target molecules in said database that bind or are modulated by said ligand, thereby determining the selectivity of said ligand of interest.

20

25

73. The method of claim 72, wherein said method is computer implemented.

30

WO 02/058533

PCT/US01/43348

74. The method of claim 72, wherein said ligand increases an activity of a target molecule, wherein said activity is associated with a disease state, an adverse side-effect, or toxicity and said ligand is eliminated from drug discovery or development or lead optimization.

5

75. The method of claim 72, wherein said ligand decreases an activity of a target molecule, wherein said activity is associated with a disease state, an adverse side-effect, or toxicity and said ligand is selected for drug discovery or development or lead optimization.

10

76. A method of for selecting a therapy for a subject for the treatment, stabilization, or prevention of a disease or disorder, said method comprising:

(a) providing an electronic database comprising at least 10 records of target molecules correlated to records of the therapeutics and their ability to bind or modulate the activity of said target molecules;

15

(b) determining a target molecule in said subject that has a mutation associated with said disease or disorder; and

(c) selecting a therapeutic from said database that binds or modulates the activity of said target molecule and thereby treats, stabilizes, or prevents said disease or disorder.

20

77. The method of claim 75, wherein said method is computer implemented.

25

WO 02/058533

PCT/US01/43348

78. A method of for selecting a therapy for a subject for the treatment, stabilization, or prevention of a disease or disorder, said method comprising:

- (a) providing an electronic database comprising at least 10 records of target molecules correlated to records of the therapeutics and their ability to bind or modulate the activity of said target molecules;
- 5 (b) determining a target molecule in said subject that has a mutation associated with said disease or disorder;
- (c) selecting a therapeutic from said database that does not bind or modulate the activity of said target molecule.

10

79. The method of claim 78, wherein said target molecule is a protein.

80. The method of claim 78, wherein said target molecule is a nucleic acid.

15

81. The method of claim 78, wherein said method is computer implemented.

82. A method of determining whether a compound of interest is present in a sample, said method comprising:

- (a) providing reference mass spectra for two or more compounds from a library of compounds;
- (b) providing a test mass spectrum of a sample comprising one or more compounds from said library; and
- 25 (c) determining whether peaks of a reference mass spectrum are included in said test mass spectrum, thereby determining whether the compound that generated said reference mass spectrum is present in said sample.

83. The method of claim 82, wherein said reference mass spectra are sequentially or simultaneously analyzed until all of the peaks in said test mass spectrum have been assigned to a compound.

30

WO 02/058533

PCT/US01/43348

84. The method of claim 82, wherein step (c) comprises a sequential determination of whether the peaks of one or more reference mass spectrum are included in said test mass spectrum.

- 5 85. The method of claim 82, wherein step (c) is repeated until either
- (i) all of the peaks in said reference mass spectrum are determined to be present in said test mass spectrum, thereby determining that the compound that generated said reference mass spectrum is present in said sample; or
 - (ii) a peak in said reference mass spectrum is determined to be absent in
- 10 said test mass spectrum, thereby determining that the compound that generated said reference mass spectrum is not present in said sample.

86. The method of claim 82, wherein step (a) comprises determining the mass spectrum of each compound in said library.

15

87. The method of claim 82, wherein at least one of the peaks in said reference spectrum is an isotope peak or a fragment peak.

88. The method of claim 82, wherein at least one of the peaks in said

20 reference spectrum is a parent peak.

89. The method of claim 82, wherein said reference mass spectrum are contained in a database comprising records of one or more properties of mass spectra correlated to references of compounds that generate said mass spectra.

25

90. The method of claim 82, wherein step (c) is computer implemented.

91. A method of determining whether a compound of interest is present in a sample, said method comprising:

- 30 (a) providing reference mass spectra for two or more compounds from a library of compounds;

WO 02/058533

PCT/US01/43348

(b) providing a test mass spectrum of a sample comprising one or more compounds from said library;

(c) determining whether one or more peaks of said test mass spectrum are included in a reference mass spectrum; and

5 (d) determining whether all of the peaks in a reference mass spectrum are present in said test mass spectrum, wherein said reference mass spectrum is a reference mass spectrum from step (c) that contains a peak present in said test mass spectrum, thereby determining whether the compound that generated said reference mass spectrum is present in said sample.

10

92. The method of claim 91, wherein step (d) comprises a sequential determination of whether the peaks of one or more reference mass spectrum are included in said test mass spectrum.

15

93. The method of claim 91, wherein step (d) comprises determining whether a peak in said reference mass spectrum is present in test mass spectrum, wherein said determination is repeated until either

(i) all of the peaks in said reference mass spectrum are determined to be present in said test mass spectrum, thereby determining that the compound that generated said reference mass spectrum is present in said sample; or

20

(ii) a peak in said reference mass spectrum is determined to be absent in said test mass spectrum, thereby determining that the compound that generated said reference mass spectrum is not present in said sample.

25

94. The method of claim 91, wherein step (a) comprises determining the mass spectrum of each compound in said library.

95. The method of claim 91, wherein at least one of the peaks in said reference spectrum is an isotope peak or a fragment peak.

30

96. The method of claim 91, wherein at least one of the peaks in said reference spectrum is a parent peak.

WO 02/058533

PCT/US91/43348

97. The method of claim 91, wherein said reference mass spectrum are contained in a database comprising records of one or more properties of mass spectra correlated to references of compounds that generate said mass spectra.

5

98. The method of claim 97, wherein said property is selected from the group consisting of: the mass to charge ratio of an isotope peak, the mass to charge ratio of a fragment peak, the mass to charge ratio of a parent peak, and the intensity of a peak.

10

99. The method of claim 97, wherein step (c) or step (d) is computer implemented.

100. A computer-readable memory having stored thereon a program for determining whether a compound of interest is present in a sample comprising:

15 a) computer code that receives as input mass spectrometry data comprising the mass to charge ratio for one or more peaks in reference mass spectra for two or more compounds from a library of compounds;

b) computer code that receives as input mass spectrometry data comprising the mass to charge ratio for one or more peaks in a test mass spectra of a sample comprising one or more compounds from said library; and

20 (c) computer code that determines whether peaks of a reference mass spectrum are included in said test mass spectrum, thereby determining whether the compound that generated said reference mass spectrum is present in said sample.

25

101. A computer-readable memory having stored thereon a program for determining whether a compound of interest is present in a sample comprising:

a) computer code that receives as input mass spectrometry data comprising the mass to charge ratio for one or more peaks in reference mass spectra for two or more compounds from a library of compounds;

30

WO 02/058533

PCT/US01/43348

b) computer code that receives as input mass spectrometry data comprising the mass to charge ratio for one or more peaks in a test mass spectra of a sample comprising one or more compounds from said library;

5 (c) computer code that determines whether one or more peaks of said test mass spectrum are included in a reference mass spectrum; and

(d) computer code that determines whether all of the peaks in a reference mass spectrum are present in said test mass spectrum, thereby determining whether the compound that generated said reference mass spectrum is present in said sample.

10

102. A method of producing two or more vectors encoding proteins of interest, said method comprising:

(a) robotically contacting a first nucleic acid encoding a first protein of interest with a first backbone nucleic acid in a first compartment in a robotic device under conditions that permit their reaction, thereby producing a first vector encoding said first protein; and

15 (b) robotically contacting a second nucleic acid encoding a second protein of interest with a second backbone nucleic acid in a second compartment in said robotic device under conditions that permit their reaction, thereby producing a second vector encoding said second protein.

20

103. The method of claim 102, further comprising:

(c) robotically contacting said first vector with a first cell under conditions that allow the insertion of said first vector into said first cell; and

25 (d) robotically contacting said second vector with a second cell under conditions that allow the insertion of said second vector into said second cell.

104. The method of claim 103, wherein said first cell expresses said first protein and said second cell expresses said second protein.

30

105. The method of claim 102, wherein at least 5 vectors are produced simultaneously

WO 02/058533

PCT/US01/43348

106. A method of purifying proteins, said method comprising:

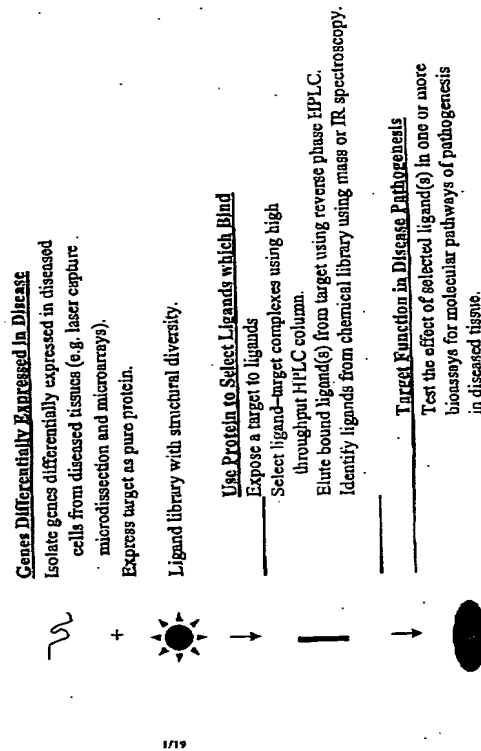
- (a) expressing a first protein in a first cell under conditions that result in the secretion of said first protein into a first medium in a robotic device;
- 5 (b) expressing a second protein in a second cell under conditions that result in the secretion of said second protein into a second medium in said robotic device;
- (c) robotically transferring said first medium to a first chromatography column and said second medium to a second chromatography column; and
- 10 (d) purifying said first protein and said second protein.

107. The method of claim 106, wherein at least 5 proteins are purified simultaneously.

WO 02/058533

PCT/US01/43348

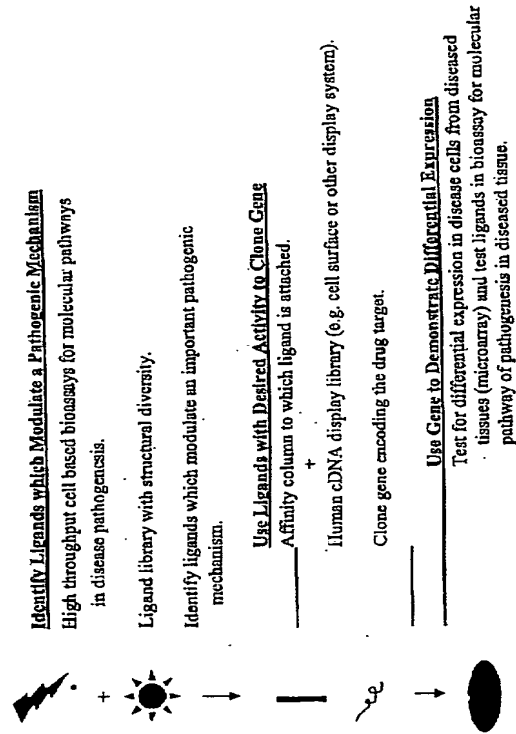
**Figure 1: Linking Genotype and Phenotype:
Overview of the Genotype to Phenotype Approach**



WO 02/059533

PCT/US01/43348

**Figure 2: Linking Genotype and Phenotype:
Overview of the Phenotype to Genotype Approach**



WO 02/058533

PCT/US91/43348

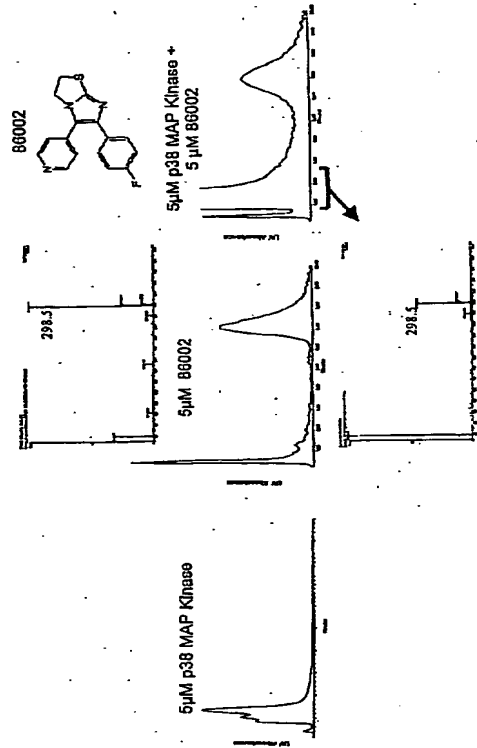
p38 MAP Kinase isolates and extracts a specific μM hit (86002).

Figure 3

WO 02/058533

PCT/US01/43348

p38 MAP Kinase binds and extracts a μ M hit (86002) but not a nonspecific control (quinine) in a concentration dependent manner.

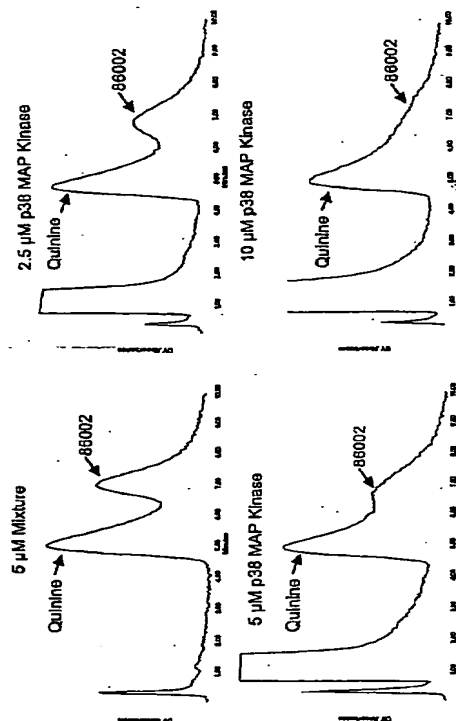
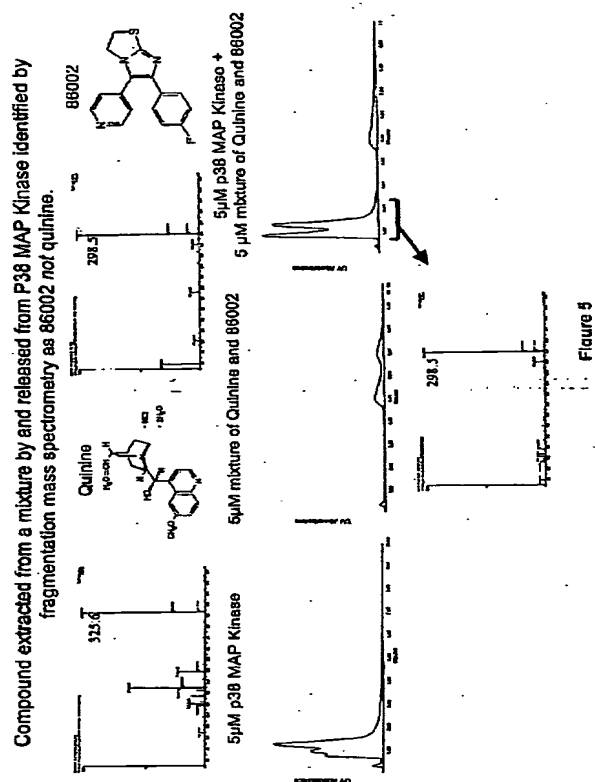


Figure 4

WO 02/058533

PCT/US01/43348



WO 02/058533

PCT/US91/43348

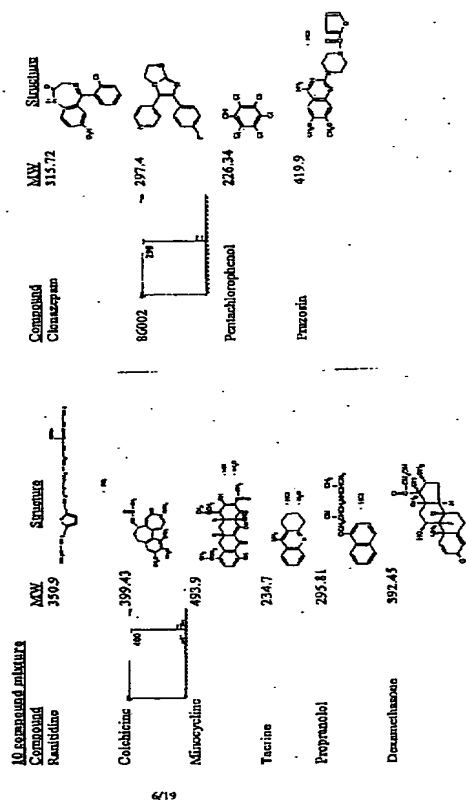
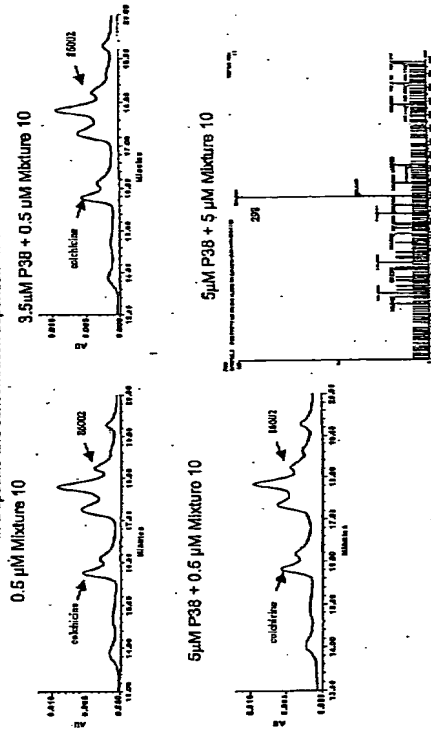


Figure 6

WO 02/058533

PCT/US91/43348

P38 MAP Kinase binds and extracts a μM hlt (8002) from a 10 compound mixture
in a specific and concentration dependent manner.



WO 02/058533

PCT/US01/43348

Tubulin binds and extracts a hit (colchicine) from a 10 compound mixture in a specific and concentration dependent manner.

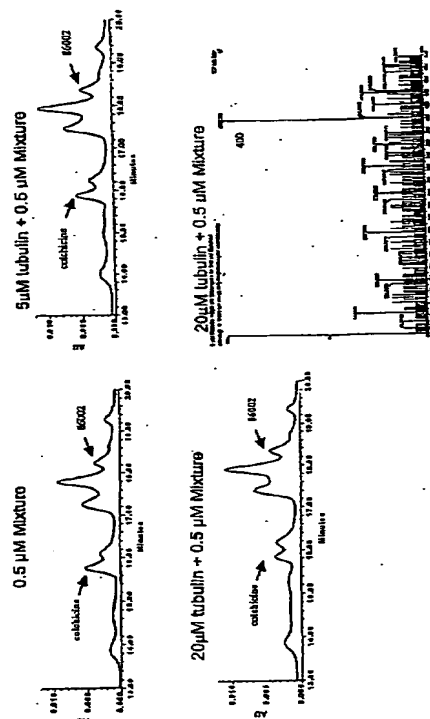


Figure 8

WO 02/058533

PCT/US01/43348

P38 MAP Kinase binds and extracts a μ M hit (88002) from a 100 compound mixture in a specific and concentration dependent manner.

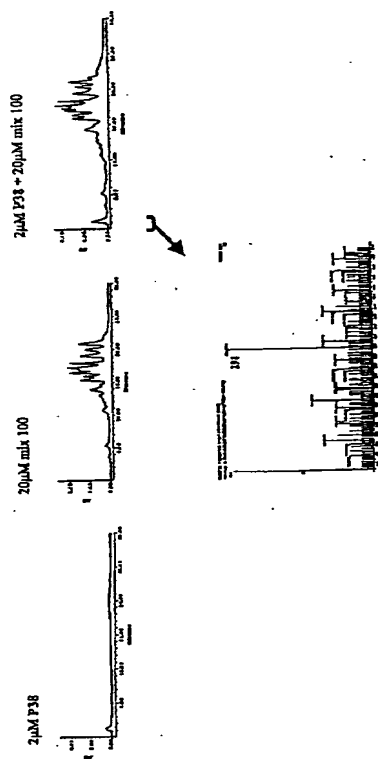


Figure 10

WO 02/058533

PCT/US01/43348

Tubulin binds and extracts a hit (colchicine) from a 100 compound mixture in a specific and concentration dependent manner.

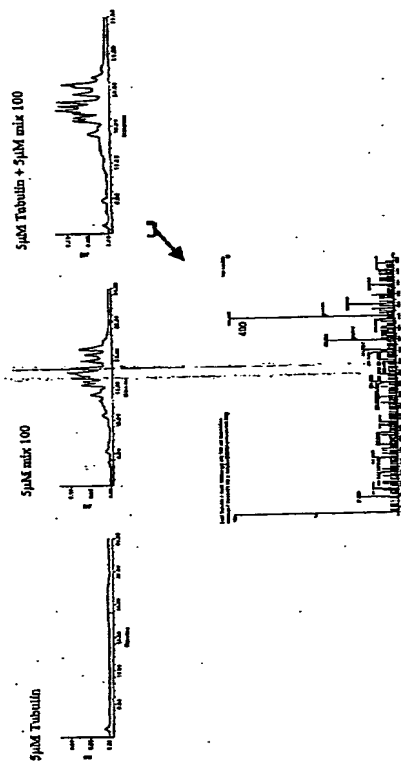


Figure 11

WO 02/058533

PCT/US01/43348

Excellent separation of protein target from 100 compound mix is also achieved at higher flow rates.

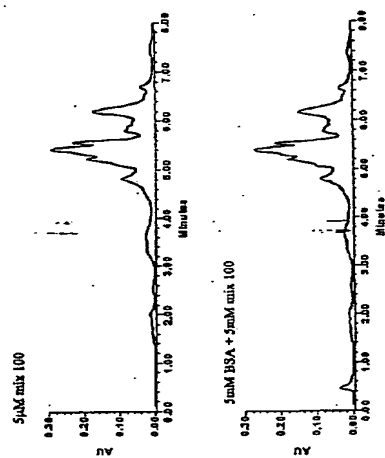
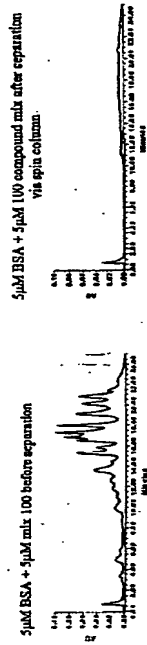


Figure 12

WO 02/058533

PCT/US01/43348

Rapid separation (<1 minute) with reasonable purity has been achieved with spin columns, a highly scalable process.



Spectra from 20 μ M Tubulin + 20 μ M mix 100 separated with spin column shows colchicine as major peak.

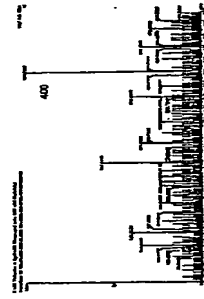


Figure 12

WO 02/058533

PCT/US01/43348

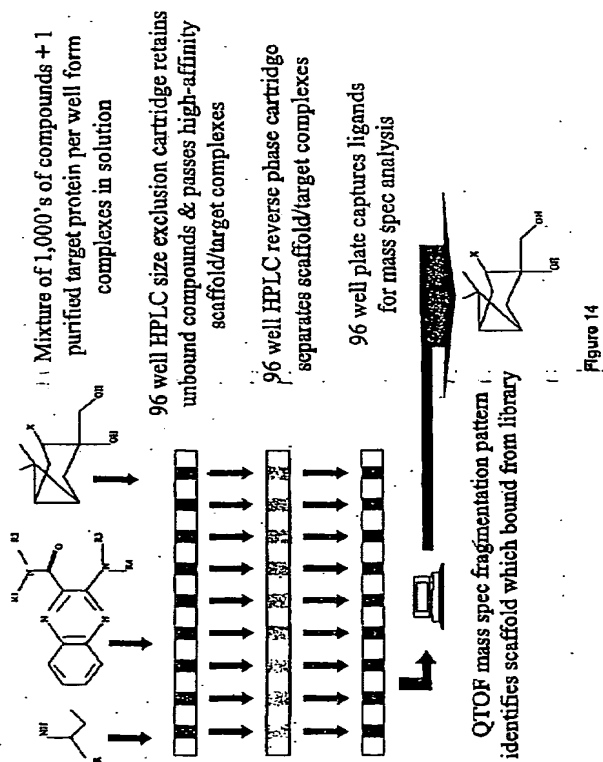


Figure 14

WO 02/058533

PCT/US01/43348

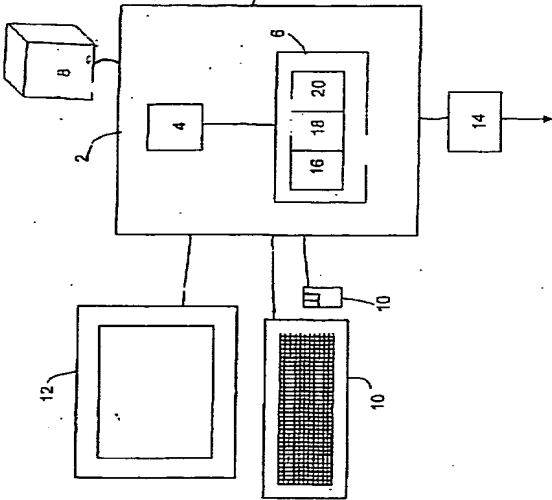


Figure 15

WO 02/058533

PCT/US01/43348

Computer Code for identifying a compound in a sample

- 1) INPUT one or more peaks in a mass spectra for each compound in a library of compounds identified
- 2) INPUT of same data for sample to be

- a. Entry of mass to charge ratio with With or without intensity for one or more peaks in spectra
- b. Entry of Digitized form of one or more peaks in spectra

Compound A	Mass/Charge A Peak1, Peak2, Peak3	
		Mass/Charge S Peak1, Peak2, Peak3
Compound B	Mass/Charge B Peak1, Peak2, Peak3	Sample

Masslynx, Oracis or Excel

- 3) Search for S Peak 1
- 4) Search for S Peak 2
- 5) Search for S Peak 3



6) For each search enter the descriptor in the compound row corresponding to the Peak which matches with that in the sample.

7) The resulting result is the compound which is present in the sample.

Figure 16

WO 02/058533

PCT/US01/43348

	Scaffold 1			Scaffold 2			Scaffold 3		
	Variant 1	Variant 2	Variant 3	Variant 1	Variant 2	Variant 3	Variant 1	Variant 2	Variant 3
Gene 1	+	+	+	+	+	+	+	+	+
Gene 2	+	+	+	+	+	+	+	+	+
Gene 3	+	+	+	+	+	+	+	+	+
Gene 4	+	+	+	+	+	+	+	+	+
Gene 5	+	+	+	+	+	+	+	+	+
Gene 6	+	+	+	+	+	+	+	+	+
Gene 7	+	+	+	+	+	+	+	+	+
Gene 8	+	+	+	+	+	+	+	+	+
Gene 9	+	+	+	+	+	+	+	+	+
Gene 10	+	+	+	+	+	+	+	+	+
Gene 11	+	+	+	+	+	+	+	+	+
Gene 12	+	+	+	+	+	+	+	+	+
Gene 13	+	+	+	+	+	+	+	+	+
Gene 14	+	+	+	+	+	+	+	+	+
Gene 15	+	+	+	+	+	+	+	+	+
Gene 16	+	+	+	+	+	+	+	+	+
Gene 17	+	+	+	+	+	+	+	+	+
Gene 18	+	+	+	+	+	+	+	+	+
Gene 19	+	+	+	+	+	+	+	+	+
Gene 20	+	+	+	+	+	+	+	+	+
Gene 21	+	+	+	+	+	+	+	+	+
Gene 22	+	+	+	+	+	+	+	+	+
Gene 23	+	+	+	+	+	+	+	+	+
Gene 24	+	+	+	+	+	+	+	+	+
Gene 25	+	+	+	+	+	+	+	+	+
Gene 26	+	+	+	+	+	+	+	+	+
Gene 27	+	+	+	+	+	+	+	+	+
Gene 28	+	+	+	+	+	+	+	+	+
Gene 29	+	+	+	+	+	+	+	+	+
Gene 30	+	+	+	+	+	+	+	+	+
Gene 31	+	+	+	+	+	+	+	+	+
Gene 32	+	+	+	+	+	+	+	+	+
Gene 33	+	+	+	+	+	+	+	+	+
Gene 34	+	+	+	+	+	+	+	+	+
Gene 35	+	+	+	+	+	+	+	+	+
Gene 36	+	+	+	+	+	+	+	+	+
Gene 37	+	+	+	+	+	+	+	+	+
Gene 38	+	+	+	+	+	+	+	+	+
Gene 39	+	+	+	+	+	+	+	+	+
Gene 40	+	+	+	+	+	+	+	+	+
Gene 41	+	+	+	+	+	+	+	+	+
Gene 42	+	+	+	+	+	+	+	+	+
Gene 43	+	+	+	+	+	+	+	+	+
Gene 44	+	+	+	+	+	+	+	+	+
Gene 45	+	+	+	+	+	+	+	+	+
Gene 46	+	+	+	+	+	+	+	+	+
Gene 47	+	+	+	+	+	+	+	+	+
Gene 48	+	+	+	+	+	+	+	+	+
Gene 49	+	+	+	+	+	+	+	+	+
Gene 50	+	+	+	+	+	+	+	+	+
Gene 51	+	+	+	+	+	+	+	+	+
Gene 52	+	+	+	+	+	+	+	+	+
Gene 53	+	+	+	+	+	+	+	+	+
Gene 54	+	+	+	+	+	+	+	+	+
Gene 55	+	+	+	+	+	+	+	+	+
Gene 56	+	+	+	+	+	+	+	+	+
Gene 57	+	+	+	+	+	+	+	+	+
Gene 58	+	+	+	+	+	+	+	+	+
Gene 59	+	+	+	+	+	+	+	+	+
Gene 60	+	+	+	+	+	+	+	+	+
Gene 61	+	+	+	+	+	+	+	+	+
Gene 62	+	+	+	+	+	+	+	+	+
Gene 63	+	+	+	+	+	+	+	+	+
Gene 64	+	+	+	+	+	+	+	+	+
Gene 65	+	+	+	+	+	+	+	+	+
Gene 66	+	+	+	+	+	+	+	+	+
Gene 67	+	+	+	+	+	+	+	+	+
Gene 68	+	+	+	+	+	+	+	+	+
Gene 69	+	+	+	+	+	+	+	+	+
Gene 70	+	+	+	+	+	+	+	+	+
Gene 71	+	+	+	+	+	+	+	+	+
Gene 72	+	+	+	+	+	+	+	+	+
Gene 73	+	+	+	+	+	+	+	+	+
Gene 74	+	+	+	+	+	+	+	+	+
Gene 75	+	+	+	+	+	+	+	+	+
Gene 76	+	+	+	+	+	+	+	+	+
Gene 77	+	+	+	+	+	+	+	+	+
Gene 78	+	+	+	+	+	+	+	+	+
Gene 79	+	+	+	+	+	+	+	+	+
Gene 80	+	+	+	+	+	+	+	+	+
Gene 81	+	+	+	+	+	+	+	+	+
Gene 82	+	+	+	+	+	+	+	+	+
Gene 83	+	+	+	+	+	+	+	+	+
Gene 84	+	+	+	+	+	+	+	+	+
Gene 85	+	+	+	+	+	+	+	+	+
Gene 86	+	+	+	+	+	+	+	+	+
Gene 87	+	+	+	+	+	+	+	+	+
Gene 88	+	+	+	+	+	+	+	+	+
Gene 89	+	+	+	+	+	+	+	+	+
Gene 90	+	+	+	+	+	+	+	+	+
Gene 91	+	+	+	+	+	+	+	+	+
Gene 92	+	+	+	+	+	+	+	+	+
Gene 93	+	+	+	+	+	+	+	+	+
Gene 94	+	+	+	+	+	+	+	+	+
Gene 95	+	+	+	+	+	+	+	+	+
Gene 96	+	+	+	+	+	+	+	+	+
Gene 97	+	+	+	+	+	+	+	+	+
Gene 98	+	+	+	+	+	+	+	+	+
Gene 99	+	+	+	+	+	+	+	+	+
Gene 100	+	+	+	+	+	+	+	+	+

Use to Determine:

- Specificity of scaffold for a target
- Potential toxicity
- Identify compound to probe n particular biology or pathology
- Select target responsible for action of a particular compound
- Select compound based on pharmacogenetics
- Select scaffolds to serve as leads for optimization of a drug

Figure 17

WO 02/058533

PCT/US01/43348

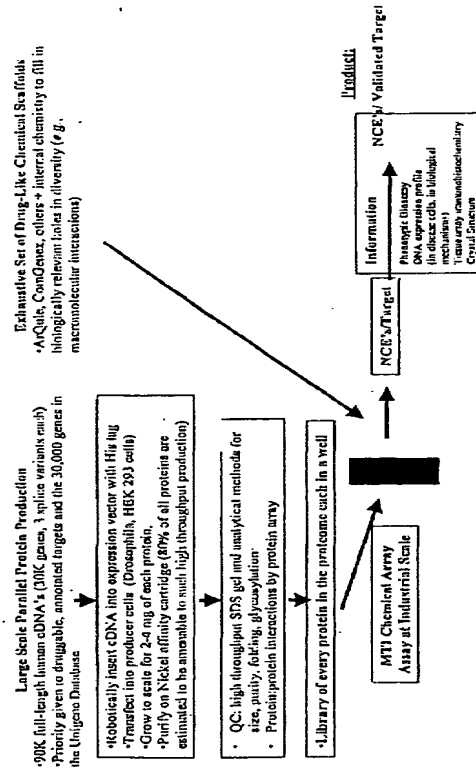
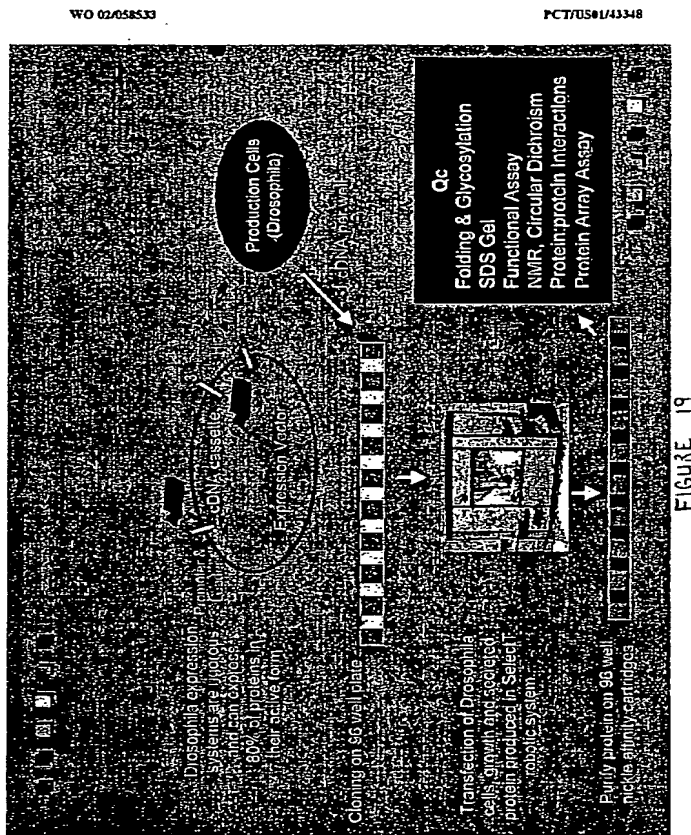


Figure 18



【国際公開パンフレット（コレクトバージョン）】

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau(43) International Publication Date
1 August 2002 (01.08.2002)

PCT

(10) International Publication Number
WO 02/058533 A3(51) International Patent Classification: G01N 33/53,
31/00

(21) International Application Number: PCT/AT01/0143348

(22) International Filing Date:
19 November 2001 (19.11.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/249,832 17 November 2000 (17.11.2000) US
60/179,463 15 October 2000 (15.10.2000) US(81) Designated States (national): AF, AG, AI, AM, AT, AU,
AZ, BA, BB, BG, BR, BY, BZ, CA, CL, CN, CO, CR, CU,
CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GR,
GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC,
LK, LR, LS, LU, LV, MA, MD, MG, MK, MN, MW,
MX, MY, NZ, NT, PT, RO, RU, SD, SE, SG, SI, SK,
SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA,
ZM, ZW.(84) Designated States (regional): AKIPO patent (GI, GM,
KE, LS, MW, NZ, SD, SL, SZ, TZ, UG, ZM, ZW),
Latin America patent (AM, AG, AR, BR, CL, CO, CR, CU,
EC, GR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC,
LK, LR, LS, LU, LV, MA, MD, MG, MK, MN, MW,
MX, MY, NZ, NT, PT, RO, RU, SD, SE, SG, SI, SK,
SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA,
ZM, ZW).(63) Related by continuation (CON) or continuation-in-part
(CIP) to earlier applications:
US 60/249,832 (CIP)
Filed on 17 November 2000 (17.11.2000)
US 60/179,463 (CIP)
Filed on 15 October 2000 (15.10.2000)Published:
with international search report(86) Date of publication of the international search report:
30 January 2003For two-letter codes and other abbreviations, refer to the "Guide-
ance Notes on Codes and Abbreviations" appearing at the begin-
ning of each regular issue of the PCT Gazette.(71) Applicant and
(72) Inventor: SLANEY, Alfred, E. [—AT]; 14 Nichols
Road, Cohasset, MA 02025 (US).(74) Agent: CLARK, Paul, E. Clark & Ebling LLP, 176 Fed-
eral Street, Boston, MA 02110-2214 (US).

WO 02/058533 A3

(54) Title: PROCESS FOR DETERMINING TARGET FUNCTION AND IDENTIFYING DRUG LEADS

(57) Abstract: The present invention relates to methods for using chemical ligands to determine target function and identify drug leads.

【国際調査報告】

INTERNATIONAL SEARCH REPORT		International application No. PCT/US01/43348
A. CLASSIFICATION OF SUBJECT MATTER IPC Cl. : G01N 31/00, 31/02 US Cl. : 435/7.1, 702/77 According to International Patent Classification (IPC) or to both national classification and IPC		
B. FIELDS SEARCHED Minimum documentation searched (classification system followed by classification symbol) U.S. : 435/7.1, 702/77 Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched Electronic data base consulted during the international search (name of data base and, where practicable, search terms used) Please See Communication Sheet		
C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category *	Character of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	DAVIE, R.G. et al. Sensitive size-exclusion chromatography coupled with liquid chromatographic mass spectrometry to identify and identify high-binding ligands from complex mixtures. Tetrahedron, 1999, Vol. 55, pages 11653-11667, especially page 11653.	1-3, 33, 34
Y	GEYKEN, R.M. et al. Isotope or mass encoding of combinatorial libraries. Chemistry and Biology August 1998, Vol. 3, pages 679-688, especially page 679.	4-6
Y	FEJTO, J. et al. The SHAFES strategy: an HPLC-based approach for lead generation in drug discovery. Chemistry and Biology, October 1999, Vol. 6, pages 755-769, especially page 755.	4-6
X		9, 10
<input type="checkbox"/> Further documents are listed in the continuation of Box C. <input type="checkbox"/> See patent family member.		
Special categories of cited documents: *A* documents defining the general state of the art which is an antecedent to the invention *B* earlier applications or prior art published on or after the international filing date *C* documents which are prior art to the invention as claimed in the claims (as specified) *D* documents relating to the invention, i.e., additions or other matters *E* documents published prior to the international filing date but later than the priority date claimed *F* documents published after the international filing date or priority date and not in accordance with the application, but which are relevant to the principles or theory underlying the invention *G* documents of particular importance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone *H* documents of particular relevance; the claimed invention cannot be considered as involving an inventive step when the document is considered with one or more other such documents, such combination being obvious to a person skilled in the art *I* documents of the same patent family		
Date of the actual completion of the international search 01 June 2002 (01.06.2002)		Date of mailing of the international search report 03 June 2002
Name and mailing address of the ISA/US Consultant of Patent Trademarks Int. PCT Washington, D.C. 20540 Facsimile No. (703)305-3220		Authorizing Officer KAROL BOEL-HARRIS JR John S. Brennan Telephone No. 703 308-0196

Form PCT/ISA/210 (second sheet) (July 1998)

INTERNATIONAL SEARCH REPORT		International application No. PCT/US01/43948
Box I Observations where certain claims were found unsearchable (Continuation of Item 1 of first sheet)		
This international report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:		
1.	<input type="checkbox"/>	Claim Nos.: because they relate to subject matter not required to be searched by this Authority, namely:
2.	<input type="checkbox"/>	Claim Nos.: because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:
3.	<input type="checkbox"/>	Claim Nos.: because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 5.4(a).
Box II Observations where unity of invention is lacking (Continuation of Item 2 of first sheet)		
This International Searching Authority found multiple inventions in this international application, as follows: Please See Continuation Sheet		
1.	<input type="checkbox"/>	As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2.	<input type="checkbox"/>	As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3.	<input type="checkbox"/>	As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claim Nos.:
4.	<input checked="" type="checkbox"/>	No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claim Nos.: 1, 10, 33, 34.
Remark on Protest	<input type="checkbox"/>	The additional search fees were accompanied by the applicant's protest.
	<input type="checkbox"/>	No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US01/03348

BOX II. OBSERVATIONS WHERE UNITY OF INVENTION IS LACKING

This application contains the following inventions or groups of inventions which are not so linked as to form a single general inventive concept under PCT Rule 13.1. In order for all inventions to be examined, the appropriate additional examination fees must be paid.

Group I, claim(s) 1-10, 33, 34, drawn to a method of screening for ligands (1st method).

Group II, claim(s) 11, 12, drawn to a method of screening for two ligands that bind simultaneously (2nd method).

Group III, claim(s) 13-20, drawn to a method of determining the function of a target (3rd method).

Group IV, claim(s) 21-32, drawn to a method of linking two ligands (4th method).

Group V, claim(s) 35-38, drawn to a method of assay of ligands that bind a fusion protein in vivo (5th method).

Group VI, claim(s) 39-42, drawn to a method of assay of ligands to target proteins by use of target proteins covalently linked to nucleic acids encoding the target proteins (6th method).

Group VII, claim(s) 43-45, drawn to a method of assay of modulators of a target (7th method).

Group VIII, claim(s) 46, 47, drawn to a method of assay of modulators of two targets (8th method).

Group IX, claim(s) 48-52, drawn to a database of ligands and target molecules (1st product; not related to methods of groups I-VIII).

Group X, claim(s) 53, 54, 58-77, drawn to a database of modulators and phenotypes and its method of use (2nd product and method of use; not related to methods of groups I-VIII).

Group XI, claim(s) 55-57, drawn to a database of expression profiles (3rd product; not related to methods of groups I-VIII).

Group XII, claim(s) 69, drawn to a method of synthesis of compounds (9th method).

Group XIII, claim(s) 76-81, drawn to a method of selecting a therapeutic regime by use of a database of modulators of phenotypes (9th method; not related to methods of groups I-VIII).

Group XIV, claim(s) 82-99, drawn to a method of assay of a compound by use of mass spectra (10th method; not related to methods of groups I-VIII).

Group XV, claim(s) 100, 101, drawn to computer readable media comprising a program for mass spectra analysis (4th product; not related to methods of groups I-VIII).

Group XVI, claim(s) 102-105, drawn to a method of robotically making expression vectors (11th method; not related to methods of groups I-VIII).

Group XVII, claim(s) 106, 107, drawn to a method of purifying two proteins (12th method; not related to methods of groups I-VIII).

The inventions listed as Groups I-XVII do not relate to a single general inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons: PCT Rule 13.1 and Annex B do not provide for unity of invention between two or more different products, methods of making, or methods of use that share a special technical feature. In addition, Groups I-VIII and XII, Group VIII, Group IX, Group X, Group XII, Group XIII, Group XIV, Group XV, and Group XVI do not share a special technical feature.

INTERNATIONAL SEARCH REPORT	International application No. PCT/US01/03453
<p data-bbox="261 548 529 590">Continuation of B. FIELDS SEARCHED Item 3: McGraw, Binda, US Patent database search terms: combinatorial, isotope, NMR, Mod, Uptid</p>	

Form PCT/ISA/210 (second sheet) (July 1998)

フロントページの続き

(51) Int. Cl. ⁷	F I	テーマコード (参考)
G 0 1 N 33/50	G 0 1 N 33/50	Z 4 H 0 4 5
G 0 1 N 33/53	G 0 1 N 33/53	D 5 B 0 7 5
G 0 1 N 33/566	G 0 1 N 33/53	M
G 0 6 F 17/30	G 0 1 N 33/566	
// A 6 1 K 31/055	G 0 6 F 17/30	1 7 0 F
A 6 1 K 31/138	A 6 1 K 31/055	
A 6 1 K 31/165	A 6 1 K 31/138	
A 6 1 K 31/341	A 6 1 K 31/165	
A 6 1 K 31/4439	A 6 1 K 31/341	
A 6 1 K 31/4709	A 6 1 K 31/4439	
A 6 1 K 31/517	A 6 1 K 31/4709	
A 6 1 K 31/5513	A 6 1 K 31/517	
A 6 1 K 31/573	A 6 1 K 31/5513	
A 6 1 K 31/65	A 6 1 K 31/573	
	A 6 1 K 31/65	

(81) 指定国 AP (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), EA (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), EP (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OA (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG), AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, R O, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW

F ターム (参考) 4B063 QA05 QQ79 QR48 QR77 QR79 QS15
 4C086 AA10 BA03 BC28 BC50 BC56 CB17 CB27 DA10 DA29
 4C206 AA10 CA18 FA19
 4H045 AA50 CA40 GA26
 5B075 UU19

**This Page is Inserted by IFW Indexing and Scanning
Operations and is not part of the Official Record.**

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☒ **BLACK BORDERS**
- ☐ **IMAGE CUT OFF AT TOP, BOTTOM OR SIDES**
- ☐ **FADED TEXT OR DRAWING**
- ☐ **BLURRED OR ILLEGIBLE TEXT OR DRAWING**
- ☐ **SKEWED/SLANTED IMAGES**
- ☒ **COLOR OR BLACK AND WHITE PHOTOGRAPHS**
- ☐ **GRAY SCALE DOCUMENTS**
- ☐ **LINES OR MARKS ON ORIGINAL DOCUMENT**
- ☐ **REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY**
- ☐ **OTHER:** _____

IMAGES ARE BEST AVAILABLE COPY.

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.